Outstanding Contributions to Logic 30

Yale Weiss
Romina Birman *Editors*

# Saul Kripke on Modal Logic

Springer

# Outstanding Contributions to Logic

Volume 30

**Editor-in-Chief**

Sven Ove Hansson, Division of Philosophy, KTH Royal Institute of Technology, Stockholm, Sweden

Outstanding Contributions to Logic puts focus on important advances in modern logical research. Each volume is devoted to a major contribution by an eminent logician. The series will cover contributions to logic broadly conceived, including philosophical and mathematical logic, logic in computer science, and the application of logic in linguistics, economics, psychology, and other specialized areas of study.

A typical volume of Outstanding Contributions to Logic contains:

- A short scientific autobiography by the logician to whom the volume is devoted
- The volume editor's introduction. This is a survey that puts the logician's contributions in context, discusses its importance and shows how it connects with related work by other scholars
- The main part of the book will consist of a series of chapters by different scholars that analyze, develop or constructively criticize the logician's work
- Response to the comments, by the logician to whom the volume is devoted
- A bibliography of the logician's publications Outstanding Contributions to Logic is published by Springer as part of the Studia Logica Library.

This book series, is also a sister series to Trends in Logic and Logic in Asia: Studia Logica Library. All books are published simultaneously in print and online. This book series is indexed in SCOPUS.

Proposals for new volumes are welcome. They should be sent to the editor-in-chief sven-ove.hansson@abe.kth.se

Yale Weiss · Romina Birman
Editors

# Saul Kripke on Modal Logic

Springer

*Editors*
Yale Weiss
The Saul Kripke Center
The Graduate Center, CUNY
New York, NY, USA

Romina Birman
The Saul Kripke Center
The Graduate Center, CUNY
New York, NY, USA

ISSN 2211-2758 ISSN 2211-2766 (electronic)
Outstanding Contributions to Logic
ISBN 978-3-031-57634-8 ISBN 978-3-031-57635-5 (eBook)
https://doi.org/10.1007/978-3-031-57635-5

© The Editor(s) (if applicable) and The Author(s), under exclusive license to Springer Nature Switzerland AG 2024

This work is subject to copyright. All rights are solely and exclusively licensed by the Publisher, whether the whole or part of the material is concerned, specifically the rights of translation, reprinting, reuse of illustrations, recitation, broadcasting, reproduction on microfilms or in any other physical way, and transmission or information storage and retrieval, electronic adaptation, computer software, or by similar or dissimilar methodology now known or hereafter developed.
The use of general descriptive names, registered names, trademarks, service marks, etc. in this publication does not imply, even in the absence of a specific statement, that such names are exempt from the relevant protective laws and regulations and therefore free for general use.
The publisher, the authors and the editors are safe to assume that the advice and information in this book are believed to be true and accurate at the date of publication. Neither the publisher nor the authors or the editors give a warranty, expressed or implied, with respect to the material contained herein or for any errors or omissions that may have been made. The publisher remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Cover Photograph by Melvin Fitting (Manchester, UK, 1969).

This Springer imprint is published by the registered company Springer Nature Switzerland AG
The registered company address is: Gewerbestrasse 11, 6330 Cham, Switzerland

If disposing of this product, please recycle the paper.

# Preface

This volume is a festschrift dedicated to Saul Aaron Kripke's contributions to modal logic. Although Kripke's work in logic, not to mention philosophy, ranged widely, it seemed appropriate to us that there should be a volume dedicated *specifically* to the field in which Kripke first made his name—as a teenager in Omaha, Nebraska—by developing his eponymous semantics for modal logic. This volume contains contributions by several of the world's most eminent logicians on an array of topics at the forefront of research in modal logic and related fields. The contributions are, however, united by a common debt to Kripke's pioneering achievements in the subject.

Work on this volume began in 2019, but its completion faced various regrettable and unforeseen obstacles (some global, others more particular). When Saul Kripke passed away suddenly on September 15, 2022, the volume was largely finished. A complete draft of his own primary contribution, "Individual Concepts: Their Logic, Philosophy, and Some of Their Uses", had just recently been sent out for refereeing; the version printed here is essentially that version, though with various editorial comments and an editorial introduction appended.

We would like to express our gratitude to the authors for the hard work they put into their contributions, and for their patience and understanding regarding the delayed completion of this volume. We are also very grateful to the referees for providing critical and valuable feedback and ensuring that the volume is of high quality. Further thanks are due to Cameron L. Johnson, for editorial support, and Sven-Ove Hansson and Springer, for helping to facilitate this project. This book, and Saul Kripke's contributions to it in particular, have been completed with support from the Saul Kripke Center at the City University of New York, Graduate Center. Finally, we are especially grateful to Saul Kripke himself for having collaborated with us and for having contributed to this volume. His work will no doubt be an inspiration for years to come.

New York, USA
December 2023

Yale Weiss
Romina Birman

# Contents


**1 Introduction** . . . . . 1
Yale Weiss and Romina Birman

**2 Saul Kripke: A Portrait of the Modal Logician as a Young Man** . . . 7
Yale Weiss and Romina Birman

**3 Beyond Knowledge of the Model** . . . . . 23
Sergei Artemov

**4 The Logic of Logical Necessity** . . . . . 43
Andrew Bacon and Kit Fine

**5 Relational Patterns, Partiality, and Set Lifting in Modal Semantics** . . . . . 93
Johan van Benthem

**6 Entailment, Mingle and Binary Accessibility** . . . . . 121
Katalin Bimbó and Jon Michael Dunn

**7 Modal, Fuzzy, ..., Vanilla Fixpoint Theories of Truth: A Uniform Approach** . . . . . 151
Melvin Fitting

**8 Logics for Rigidity** . . . . . 193
James W. Garson

**9 A Letter from Kripke to Lewis** . . . . . 209
Saul A. Kripke

**10 Individual Concepts: Their Logic, Philosophy, and Some of Their Uses** . . . . . 213
Saul A. Kripke

**11 Nested Sequents or Tree-Hypersequents—A Survey** . . . . . 243
Björn Lellmann and Francesca Poggiolesi

**12 A Proof-Theoretic Approach to Formal Epistemology** . . . . . 303
Sara Negri and Edi Pavlović

**13 Mission Impossible** . . . . . 347
Graham Priest

**14 Substructural Negations as Normal Modal Operators** . . . . . 365
Heinrich Wansing

**15 New(ish) Foundations for Theories of Entailment** . . . . . 389
Yale Weiss

**16 Accepting a Logic, Accepting a Theory** . . . . . 409
Timothy Williamson

**Appendix A: Saul Kripke: Complete Bibliography** . . . . . 435

# Editors and Contributors

## About the Editors

**Yale Weiss** is the Assistant Director of the Saul Kripke Center. He received his Ph.D. from the Graduate Center, CUNY in 2019. He has published widely on logic (including on conditional, connexive, intuitionistic, modal, and relevance logic) and its history (with a focus on ancient logic).

**Romina Birman** (née Padró) is the Director of the Saul Kripke Center and Saul Kripke's literary executor. She received her Ph.D. from The Graduate Center, CUNY in 2015, with a dissertation entitled, "What the Tortoise Said to Kripke: The Adoption Problem and the Epistemology of Logic". Her research focuses on the epistemology of logic and the philosophy of language.

## About the Contributors

**Sergei Artemov** is a Distinguished Professor of computer science, mathematics, and philosophy at the Graduate Center of the City University of New York. His research interests include proof theory, logic in computer science, epistemology, and epistemic game theory. He is best known for his invention of the logic of proofs and justifications. From 1996 to 2001, he was a professor of mathematics and computer science at Cornell University. He also held visiting faculty positions at Stanford University, the Universities of Amsterdam, Bern, and Siena, and other academic centers. He graduated from Moscow University with honors; his mentor there was Andrei Kolmogorov. Artemov has been an editor for several leading outlets in logic, including the *Annals of Pure and Applied Logic* and the monograph series *Studies in Logic and Foundations of Mathematics.*

**Andrew Bacon** is a professor at the University of Southern California. He has written books and articles on various topics in logic, metaphysics, and epistemology, including vagueness, higher-order logic in metaphysics, conditionals, and the liar paradox.

**Johan van Benthem** is a University Professor Emeritus of Logic at the University of Amsterdam, Henry Waldgrave Stuart Professor of Philosophy at Stanford University, and Jin Yuelin Professor of Logic at Tsinghua University. He has published on pure and applied modal logic, natural language semantics, interfaces of logic and philosophy, and the foundations of computing, information dynamics, and games. He is a member of the Dutch Academy of Sciences, the Academia Europaea, and a foreign member of the American Academy of Arts and Sciences.

**Katalin Bimbó** is a professor at the University of Alberta in Edmonton, Canada. She received her Ph.D. at Indiana University in Bloomington (1999), with J. M. Dunn as her thesis supervisor; then she held academic positions in New Zealand, Australia, and the USA. Her research focuses on non-classical logic, primarily, on relevance logic, substructural logic, and combinatory logic. She is an author of three books and numerous articles. She is an editor of the *Bulletin of Symbolic Logic*. Bimbó has received multiple federal research grants. Her research results include Church–Rosser-type theorems for extended combinatory logic, full topological duality for ortholattices, an ideal-based gaggle semantics for relevance logics, and sequent calculi for the positive logics of entailment and ticket entailment. Bimbó and Dunn proved both implicational ticket entailment and full linear logic decidable, and they worked together on many other projects.

**J. Michael Dunn** (June 19, 1941–April 5, 2021) worked on information-based logic, algebraic logic, proof theory, relevance logic, quantum logic, and logic in computer science (including quantum computers). He received an A.B. in Philosophy from Oberlin College (1963) and a Ph.D. in Philosophy (Logic) from the University of Pittsburgh (1966). His career started at Wayne State University in 1966. He retired from Indiana University in Bloomington (2007) after 38 years there. He was Oscar Ewing Emeritus Professor of Philosophy, Emeritus Professor of Computer Science and Informatics, and was Founding Dean of the School of Informatics. He was an affiliate of the Info-Metrics Institute at the American University. Dunn published 8 books and over 100 articles. He was an Editor of the *Journal of Symbolic Logic* and Editor-in-Chief of the *Journal of Philosophical Logic*. Dunn received many honors and was a Fellow of the American Academy of Arts and Sciences.

**Kit Fine** is University Professor and Silver Professor of Philosophy and Mathematics at New York University. His research interests center on metaphysics, logic, and the philosophy of language, and he has also done work in linguistics, economic theory, and computer science.

**Melvin Fitting** was born in Troy, New York, in 1942. His undergraduate degree was from Rensselaer Polytechnic Institute, in mathematics, and his 1968 Ph.D. was supervised by Raymond Smullyan, at Yeshiva University. His dissertation became

his first book, *Intuitionistic Logic, Model Theory, and Forcing* (1969). He has worked in areas including modal logic, semantics for logic programming, theory of truth, justification logic, and tableau theorem proving. In 2012 he received the Herbrand Award from the Conference on Automated Deduction and in 2019 an honorary doctorate from the University of Bucharest. He taught at City University of New York from 1969 to his retirement in 2013, at the undergraduate Lehman College, and at the Graduate Center, where he was in the Departments of Mathematics, Computer Science, and Philosophy. He has authored or co-authored a number of books and research papers.

**James W. Garson** is a Professor in the Department of Philosophy at the University of Houston. His dissertation on temporal logic was supervised by Nuel Belnap. He has taught at the University of Pennsylvania, the University of Notre Dame, Rice University, and the University of Illinois at Chicago where he was a visiting professor in Computer Science. His research interests include modal logic, the philosophy of mind, neural networks, formal semantics, and natural language processing. His publications include articles on logic, semantics, linguistics, the philosophy of cognitive science, and computerized education. He is the author of *Modal Logic for Philosophers* (Cambridge 2013), which focuses on quantified modal logic, and *What Logics Mean* (Cambridge 2013), which is a formal study of the prospects for an inferential theory of meaning for the logical connectives.

**Saul Aaron Kripke** (November 13, 1940–September 15, 2022), who made significant contributions to the model theory of modal logic while still in High School, was one of the most distinguished logicians and philosophers of his time. Besides his many contributions to mathematical and philosophical logic, he is known for his landmark philosophical books *Naming and Necessity* and *Wittgenstein on Rules and Private Language*, as well as a number of influential articles.

**Björn Lellmann** is currently working in the Federal Ministry of Finance Austria, applying logical methods in Austrian e-government. He received his Diploma in mathematics from Albert-Ludwigs Universität Freiburg. His dissertation *Sequent Calculi with Context Restrictions and Applications to Conditional Logics* (Imperial College London, 2013) was supervised by Dirk Pattinson and Marek Sergot. He worked at Technical University Vienna from 2013 to 2021. His research interests include general methods in proof theory and modal logic as well as applications of these methods for ethical AI.

**Sara Negri** is Professor of Mathematical Logic at the University of Genoa. After a Ph.D. in Mathematics in 1996 at the University of Padova and research visits at the University of Amsterdam and Chalmers, she has been a research associate at the Imperial College in London, a Humboldt Fellow in Munich, and a visiting scientist at the Mittag-Leffler Institute in Stockholm and at the Hausdorff Institute for Mathematics in Bonn. She was appointed Docent of Logic at the University of Helsinki in 1998 and became a full professor of Theoretical Philosophy in 2015. Her research interests range from mathematical logic and philosophy of mathematics to proof theory and its applications to modal and non-classical logic.

**Edi Pavlović** was born in Rijeka, Croatia. He received his bachelor's degree at the University of Rijeka and earned his master's degree in 2012 and a Ph.D. in 2017, both in philosophy, at the Central European University. He joined University of Helsinki in 2018 as a postdoctoral researcher working with Sara Negri, and since 2021 he has worked as a postdoctoral researcher at the Munich Center for Mathematical Philosophy (MCMP) with Norbert Gratzl. His research interests are primarily in proof theory and philosophical logic, in particular deontic, epistemic, and logic of agency, as well as theories of quantification.

**Francesca Poggiolesi** received her BA in Philosophy from the University of Florence, and her Ph.D. in philosophy from the University of Florence and the University of Paris 1 Panthéon-Sorbonne. Since 2012, she has been a CNRS researcher at the IHPST, UMR 8590, Paris. She is the author of the book *Gentzen Calculi for Modal Propositional Logic*. After several years working on the proof theory of non-classical logic, in particular modal logic, she has more recently turned her interests toward the notion of grounding and explanation from a logical perspective. She is one of the associated editors of the *Journal of Philosophical Logic*.

**Graham Priest** has held chairs of philosophy in Australia, the UK, and the United States, as well as many visiting positions at universities in countries in Europe and Asia. He is currently a Distinguished Professor of Philosophy at the Graduate Center, City University of New York, and Boyce Gibson Professor Emeritus at the University of Melbourne. He is known for his work on metaphysics and non-classical logic, on the history of philosophy, and on Buddhist philosophy. He has published nine monographs and over 300 articles.

**Heinrich Wansing** is a professor of Logic and Epistemology at the Ruhr-University Bochum (Germany). Before that he was a professor of Philosophy of Science and Logic at TU Dresden (1999–2010). He took his M.A. and his Ph.D. in philosophy at the Free University of Berlin and his Habilitation in logic and analytical philosophy at the University of Leipzig. He is the author of *The Logic of Information Structures* (1993), *Displaying Modal Logic* (1998), *Truth and Falsehood: An Inquiry into Generalized Logical Values* (with Y. Shramko, 2011), *Proof-Theory of N4-related Paraconsistent Logics* (with N. Kamide, 2015), and numerous articles. Heinrich Wansing has been working mainly on philosophical logic. Moreover, he is the editor-in-chief of the book series *Trends in Logic* (Springer), a managing editor of the journal *Studia Logica*, and a member of a number of other editorial boards of logic and philosophy journals.

**Timothy Williamson** became the Wykeham Professor of Logic at Oxford University in 2000. He has also been the annual Whitney Griswold Visiting Professor at Yale since 2018. His books include *Identity and Discrimination*, *Vagueness*, *Knowledge and its Limits*, *The Philosophy of Philosophy*, *Modal Logic as Metaphysics*, *Tetralogue*, *Doing Philosophy: From Common Curiosity to Logical Reasoning*, *Suppose and Tell: The Semantics and Heuristics of Conditionals*, and *Overfitting and Heuristics in Philosophy*. There are edited collections of essays on his work with his responses: *Williamson on Knowledge* and *Williamson on Modality*. He is Fellow of

the British Academy, Fellow of the Royal Society of Edinburgh, Foreign Member of the Norwegian Academy of Science and Letters, International Member of the American Academy of Arts and Sciences, Member of the Academia Europaea, Honorary Member of the Royal Irish Academy, and Member of the Institut International de Philosophie. He gave the 2016 Saul Kripke Lecture.

# Chapter 1
# Introduction

**Yale Weiss and Romina Birman**

**Abstract** This chapter serves as a brief introduction to the contents of this multi-author volume. We situate and survey the contributions and note (as applicable) the relevance of Kripke's work to them.

**Keywords** History of logic · Kripke's contributions to logic · Modal logic · New work in modal logic · Saul Kripke

Saul Aaron Kripke (November 13, 1940–September 15, 2022) made numerous contributions to logic and philosophy. This volume is dedicated to examining just one slice of his total intellectual output, albeit an important one: his contributions to modal logic.[1] In what follows, we briefly survey the contents of this volume and discuss connections to Kripke's work as appropriate.

In "Saul Kripke: A Portrait of the Modal Logician as a Young Man", we (the editors) tell the story of how Kripke first became interested in logic and modal logic in particular. We situate Kripke's achievements in the field, many of which were made as an adolescent, against the state of the discipline before his involvement. We examine several of his specific innovations in modal model theory as well as the roles of the figures who most influenced him (among them, Curry and Prior). Our narrative emphasizes the themes of historical contingency and philosophically-motivated technical work.

Kripke's semantics has been applied to many different interpreted modal logics, among them, deontic logic, provability logic, tense logic, and epistemic logic. In "Beyond Knowledge of the Model", Sergei Artemov examines *epistemic structures*,

[1] Nevertheless, many contributions to this volume touch upon other areas of Kripke's work, as we note below.

Y. Weiss (✉) · R. Birman
The Saul Kripke Center, The Graduate Center, CUNY, 365 5th Avenue, Room 7118, New York, NY 10016, USA
e-mail: yweiss@gradcenter.cuny.edu

R. Birman
e-mail: rpadro@gc.cuny.edu

© The Author(s), under exclusive license to Springer Nature Switzerland AG 2024

Y. Weiss and R. Birman (eds.), *Saul Kripke on Modal Logic*, Outstanding Contributions to Logic 30, https://doi.org/10.1007/978-3-031-57635-5_1

a sort of generalization of Kripke models intended to be able to model epistemic situations beyond their reach, as well as *observable models*. Observable models do not alter the logic, but Artemov contends that they have foundational value in that they allow modeling epistemic situations without necessarily meeting, as Kripke models must, the restrictive "common knowledge of the model" constraint.

As we remark in "Saul Kripke: A Portrait of the Modal Logician as a Young Man", one can distinguish three strands in the early to mid-20th century development of modal semantics: the algebraic, substitutional, and possible-worlds approaches. Andrew Bacon and Kit Fine revisit the substitutional approach, which was pioneered by McKinsey (1945), in their contribution, "The Logic of Logical Necessity". Roughly, the idea is that for 'It is necessary that $\varphi$' to be true is for $i(\varphi)$ to be true, for every substitution $i$ from some suitably restricted class of substitutions. Among other things, Bacon and Fine discuss the relation of this semantics to Kripke's, consider the logical interpretation of $\Box$, and examine how to extend the substitutional approach to modal predicate logic. Along the way, they solve various technical problems and present new proofs concerning two old problems of Friedman (1975).

In "Relational Patterns, Partiality, and Set Lifting in Modal Semantcs", Johan van Benthem surveys the theme of relational structure between worlds as the essential innovation of modal semantics. This relational understanding of modality is illustrated through various highlights from history of creating new ordered structures and matching modal languages, including set lifting constructions and recent hyperintensional semantic frameworks. In a similar broad vein, van Benthem's piece takes a relational view of the realm of modal systems as connected by various translations, and points out new perspectives, philosophical issues, and avenues for technical research.

Among Kripke's lesser known contributions to logic are his early results in relevance logic, including his refutation of Belnap's conjecture that the pure implicational fragment of **E** coincides with the intersection of the pure implicational fragments of **R** and **S4** (considered qua system of strict implication).[2] Kripke's refutation inspired modified conjectures, in which **S4** is replaced by one of two "mingled" versions of the logic **E**. In "Entailment, Mingle and Binary Accessibility", Katalin Bimbó and the late J. Michael Dunn refute *both* modified conjectures, in one case drawing on Kripke's own counterexample to the original conjecture. The authors also give a 3-valued Kripke-style binary relational semantics for the logic of entailment with mingle (**EM**).

The technical work Kripke is best known for besides his work on modal logic is likely his theory of truth, as outlined in Kripke (1975). In "Modal, Fuzzy, …, Vanilla Fixpoint Theories of Truth: A Uniform Approach", Melvin Fitting examines fixpoint constructions in theories of truth from the general setting of *bilattices*. Roughly, a bilattice consists of a set with two interconnected orders on it, each of which is a lattice. Fitting observes how valuations in various formal languages, including modal languages, naturally give rise to bilattices. Fitting then goes on to survey a number of quite general fixpoint theorems.

[2] Kripke's counterexample is reproduced in Anderson and Belnap (1975, Sect. 8.10, p. 94).

Kripke (1972/1980) famously argued that proper names are *rigid designators* and, consequently, where $a$ and $b$ are names, $a = b$ is necessary if true at all. In "Logics for Rigidity", James Garson examines ways of semantically regimenting the concept of a rigid designator which are both faithful to Kripke's intentions and bear out the aforementioned necessary-if-true-at-all feature of identities. As he shows, this is not as straightforward as it might have seemed. Garson focuses on three precisifications of the concept—strong rigidity, existential rigidity, and referential rigidity—and shows that each either fails to deliver the desired necessity or comes with considerable theoretical costs.

This volume contains two previously unpublished pieces on modal logic by Saul Kripke. The first, "A Letter from Kripke to Lewis", is just that: a letter from Kripke to David Lewis, dated August 11, 1969, concerning the latter's *counterpart theory*. Lewis (1968) had developed the view that individuals are world-bound, but have counterparts that stand-in for them in other worlds in virtue of which we can make nontrivial modal claims about them. In this letter, Kripke advances several important philosophical and technical criticisms of Lewis's proposal from that article.[3]

The second work of Kripke's occurring in this volume for the first time is his "Individual Concepts: Their Logic, Philosophy, and Some of Their Uses".[4] This paper, which was among the last things Kripke worked on prior to his death, contains philosophical reflections and technical results on modal logic with quantification over *individual concepts* (functions from worlds to individuals); such an approach to quantified modal logic had been advocated for by Carnap (1947/1956). Among other things, Kripke sketches an argument to the effect that many modal logics with quantification over (all) individual concepts are not axiomatizable.

Although Kripke is best known for his *semantic* (i.e., model-theoretic) work on logic, it is worth remembering that, particularly in his youth, he did a considerable amount of proof-theoretic work (consult the biographical chapter for some examples). So it is fitting that this volume should contain contributions dealing with the state-of-the-art in modal proof-theory.

Björn Lellmann and Francesca Poggiolesi survey a variation on conventional Gentzen systems in "Nested Sequents or Tree-Hypersequents—A Survey". Lellmann and Poggiolesi discuss how difficulties encountered in obtaining proof-theoretically well-behaved conventional Gentzen calculi for certain modal logics provided the impetus for proof-theoretic innovations, including new *nested sequent*, *deep sequent*, or *tree-hypersequent* calculi. The basic idea is to generalize the notion of a sequent to that of a nested sequent, where this is simply a collection of sequents arranged tree-wise. The authors sketch a number of technical results and applications concerning such calculi with a focus on various modal logics. They also exhaustively survey the connections of the nested sequent proof-theoretic framework to other proof-theoretic frameworks.

---

[3] Some of Kripke's observations from the letter are discussed in Dorr et al. (2021, Sect. 10.2). See also Kripke (1972/1980, p. 45, n. 13).

[4] An extended abstract was previously published as Kripke (1992).

In "A Proof-Theoretic Approach to Formal Epistemology", Sara Negri and Edi Pavlović examine epistemic notions such as (simple) belief and knowledge from the perspective of Conditional Doxastic Logic (**CDL**), a species of dynamic epistemic logic, using the tools of labelled sequent calculi. Labelled sequent calculi, unlike conventional Gentzen systems or certain other variations thereof, *internalize* semantic information, but can still be readily used for various purposes for which standard sequent systems are used (e.g., giving decision procedures). After developing their proof-theoretic tools, Negri and Pavlović apply them to "paradox" analysis and control, for example, in analyzing what goes amiss in cases such as those of Gettier (1963).

Kripke (1965a, pp. 210–211) introduced the concept of a *non-normal world*, a world where everything is possible and nothing is necessary, in order to model non-normal modal logics such as the Lewis systems **S2** and **S3**. In retrospect, some have viewed this as the invention of so-called *impossible worlds*, which have since come to be seen by many as a valuable theoretical device for such purposes as modeling counterpossibles and the contents of mental states. In "Mission Impossible", Graham Priest seeks to give a general characterization of the notion of an impossible world which is independent of particular applications and remains neutral on fundamental questions such as what inferences or logical laws are valid. His basic guiding principle is that while possible worlds are those closed under logical consequence, impossible worlds are not. Priest considers different ways of articulating this principle as well as various related issues.

Kripke's seminal work on modal logic focused on positive modalities, such as necessity, though he effectively also modeled negation as impossibility in his semantics for intuitionistic logic (Kripke, 1965b). In "Substructural Negations as Normal Modal Operators", Heinrich Wansing examines a theory of negations as impossibility and as unnecessity developed by Onishi (2015). Onishi had proposed the identification of these two notions of negation as a solution to a certain proof-theoretic problem; Wansing develops a different proposal in his contribution by reformulating Onishi's display calculus.

An important syntactic insight in modal logic predating Kripke's work was the realization that many logics of strict implication—notably, the Lewis system **S4**—could be presented in such a way as to separate the modal and non-modal elements of the theory. Thus, as Gödel (1933/1986) observed, **S4** could be axiomatized by extending classical propositional logic with a few axioms and rules specific to handling necessity. In "New(ish) Foundations for Theories of Entailment", Yale Weiss examines theories of *entailment* (i.e., relevant strict implication) as developed in a similar fashion (i.e., as relevant modal theories). Admissibility, semantic characterization, translation, and other results are proved.

Putnam (1968) suggested that a new logic, quantum logic, be adopted in the light of scientific discoveries in quantum mechanics. Kripke (2024) argues that this notion of "adopting a logic" is deeply confused. In "Accepting a Logic, Accepting a Theory", Timothy Williamson responds to Kripke's critique and articulates a view on which accepting a logic is akin to accepting a novel scientific theory. Williamson

illustrates the approach by examining the case of theory choice among competing interpreted systems of quantified modal logic.

## References


Anderson, A. R., & Belnap, N. D. (1975). *Entailment: The logic of relevance and necessity*. Princeton: Princeton University Press.

Carnap, R. (1947/1956). *Meaning and necessity: A study in semantics and modal logic*. Chicago: Phoenix Books (University of Chicago Press).

Dorr, C., Hawthorne, J., & Yli-Vakkuri, J. (2021). *The bounds of possibility: Puzzles of modal variation*. Oxford: Oxford University Press.

Friedman, H. (1975). One hundred and two problems in mathematical logic. *Journal of Symbolic Logic, 40*(2), 113–129.

Gettier, E. L. (1963). Is justified true belief knowledge? *Analysis, 23*(6), 121–123.

Gödel, K. (1933/1986). Eine interpretation des intuitionistischen aussagenkalküls. In S. Feferman, J. W. Dawson, Jr., S. C. Kleene, G. H. Moore, R. M. Solovay, & J. van Heijenoort (Eds.), *Kurt Gödel: Collected Works* (Vol. 1, pp. 300–303). New York: Oxford University Press.

Kripke, S. A. (1965a). Semantical analysis of modal logic II. Non-normal modal propositional calculi. In J. W. Addison, L. Henkin, & A. Tarski (Eds.), *The Theory of Models, Proceedings of the 1963 International Symposium at Berkeley* (pp. 206–220). Amsterdam: North Holland Publishing Company.

Kripke, S. A. (1965b). Semantical analysis of intuitionistic logic I. In J. N. Crossley & M. A. E. Dummett (Eds.), *Formal Systems and Recursive Functions, Proceedings of the Eighth Logic Colloquium At Oxford, July, 1963* (pp. 92–130). Amsterdam: North Holland Publishing Company.

Kripke, S. A. (1972/1980). *Naming and necessity*. Cambridge: Harvard University Press.

Kripke, S. A. (1975). Outline of a theory of truth. *Journal of Philosophy, 72*(19), 690–716.

Kripke, S. A. (1992). Individual concepts: Their logic, philosophy, and some of their uses (abstract). *Proceedings and Addresses of the American Philosophical Association, 66*(2), 70–73.

Kripke, S. A. (2024). The question of logic. *Mind, 133*(529), 1–36.

Lewis, D. (1968). Counterpart theory and quantified modal logic. *Journal of Philosophy, 65*(5), 113–126.

McKinsey, J. C. C. (1945). On the syntactical construction of systems of modal logic. *Journal of Symbolic Logic, 10*(3), 83–94.

Onishi, T. (2015). Substructural negations. *Australasian Journal of Logic, 12*(4), 177–203.

Putnam, H. (1968). Is logic empirical? In R. S. Cohen & M. W. Wartofsky (Eds.), *Boston studies in the philosophy of science* (Vol. 5, pp. 216–241). Dordrecht: D. Reidel Publishing Company.

# Chapter 2
# Saul Kripke: A Portrait of the Modal Logician as a Young Man

**Yale Weiss and Romina Birman**

**Abstract** In this short intellectual biography, we chronicle Saul Kripke's involvement in the development of modal logic, focusing on the decade beginning in 1953 and ending in 1963, during which time he ranged in age from 12 to 23. We also describe the state of modal logic before Kripke, Kripke's correspondence with other modal logicians, and Kripke's early influential publications on the semantics of modal logic as well as several later and lesser known contributions.

**Keywords** Biography · Development of modal logic · History of Kripke models · Kripke's contributions to modal logic · Kripke's correspondence · Modal logic

## 2.1 Introduction

This is a volume on Saul Aaron Kripke's (November 13, 1940–September 15, 2022) contributions to *modal logic* and, in this biographical sketch, we shall accordingly attempt to restrict our attention to those parts of his biography that are germane to his contributions to that subject.[1] It was, of course, mainly his work on modal logic that first brought Kripke to the attention of the international logic and philosophy community and most of the work he did on the subject was already completed in his

[1] Although it is customary in the series *Outstanding Contributions to Logic* for the subject logician to write their own intellectual biography, Saul Kripke had not yet begun this at the time of his death. However, he had worked closely with Romina Birman (née Padró) on an unpublished presentation, "Mrs. Croft and the History of Modal Logic", which describes how he first became interested in the subject and started to work on it. We have made extensive use of this document in preparing this biographical sketch (in particular, in writing Sects. 2.3 and 2.4).

Y. Weiss (✉) · R. Birman
The Saul Kripke Center, The Graduate Center, CUNY, 365 5th Avenue, Room 7118, New York, NY 10016, USA
e-mail: yweiss@gradcenter.cuny.edu

R. Birman
e-mail: rpadro@gc.cuny.edu

© The Author(s), under exclusive license to Springer Nature Switzerland AG 2024

Y. Weiss and R. Birman (eds.), *Saul Kripke on Modal Logic*, Outstanding Contributions to Logic 30, https://doi.org/10.1007/978-3-031-57635-5_2

youth (indeed, his adolescence). Thus, our narrative will largely focus on the decade beginning in 1953 and ending in 1963, during which time Kripke ranged in age from 12 to 23.

The paper is organized as follows. In Sect. 2.2, we briefly recount the state of modal logic before Kripke (or rather, before Kripke's involvement therein). In Sect. 2.3, we describe how Kripke came to be interested in logic and modal logic in particular. In Sect. 2.4, we discuss correspondence between Kripke and Curry, Prior, and others, in which he made many of his first important technical observations on modal logic. In Sect. 2.5, we discuss Kripke's publications on modal logic, both epoch making and lesser known. Some concluding remarks are offered in Sect. 2.6.

## 2.2 Modal Logic Before Kripke

The history of formal modal logic is as old as the history of formal logic: Aristotle develops both his theory of the assertoric and modal syllogistic in the same work, the *Prior Analytics*.[2] Its main champion in the early 20th century was C. I. Lewis, who was primarily interested in necessitated, or *strict*, conditionals. Lewis and Langford (1932/1959) presented an array of modal logics, including **S4** and **S5**, axiomatically; their original axiomatizations were rather complicated, but they would ultimately be considerably simplified by Gödel (1933/1986), Lemmon (1957), and others. These systems were originally presented and studied without any accompanying *semantics*, or *model theory*.

One might distinguish three strands in the early 20th century development of semantics for modal logic. Chronologically, the first of these was the algebraic approach.[3] The algebraic approach yielded a number of important results, including the decidability of **S2** and **S4** (McKinsey, 1941), a proof that the translation of intuitionistic logic into **S4** proposed by Gödel (1933/1986) is exact (McKinsey & Tarski, 1948), and that **S5** has no finite characteristic matrix but that each of its

[2] While the modal syllogistic never exerted quite the same level of influence as the assertoric syllogic, and interpreters have tended to regard it as confused—see, for example, Kneale and Kneale (1962, pp. 86ff.); consult Malink (2013) for a more recent (and equitable) assessment—modal logic seems to have nevertheless always attracted at least some level of interest among at least some logicians from Aristotle onwards. Indeed, the renaissance of interest in modal logic in the contemporary period, stimulated such as it was by attempts to find a superior account of the conditional freed of the so-called paradoxes of material implication, is interestingly prefigured by a debate about conditionals dating from the Hellenistic period and recounted by Sextus (*PH* 2.110ff.).

[3] It is not entirely clear who should be credited as the 'originating figure' for this approach. Łukasiewicz (1920/1970, 1930/1970) developed a three-valued logic with a modal interpretation of the third value by 1920, but this has little connection to the Lewis systems or later algebraic work on them. Many of the matrix-based consistency and independence results concerning the Lewis systems found in Lewis and Langford (1932/1959, Appendix II) are attributed to Wajsberg or Parry; see also Parry (1934). These results are, however, rather elementary and unsystematic compared with the later work of, say, McKinsey.

proper normal extensions do (Dugundji, 1940; Scroggs, 1951).[4] The mathematical fecundity and richness of the algebraic approach notwithstanding, it did not provide a semantics in the usual sense—that is, it did not articulate an intuitive framework for interpreting $\Box$ and $\Diamond$, but treated these as mere mathematical constructs (ultimately, as operations in a Boolean algebra with operators, or BAO).[5]

The second strand in the early development of semantics for modal logic was the substitutional approach pioneered by McKinsey (1945).[6] The intuitive idea is conveyed clearly in his own words:

> I take the position that to say a sentence is possible means that there exists a true sentence of the same form. Thus, for example, it would be said that the sentence, "Lions are indigenous to Alaska," is possible, because of the fact that the sentence, "Lions are indigenous to Africa," has the same form and is true. (McKinsey, 1945, p. 83)

More abstractly, fixing a set of substitutions, a sentence is called possible "if some substitution of the set takes it into a true sentence" (McKinsey, 1945, p. 83). This idea would later be further developed by Smiley (1956) and Drake (1962).[7] Although the idea is, in a broad sense, semantic—a truth condition is given for $\Diamond$-expressions and, derivatively, $\Box$-expressions—it has an undeniably syntactic aura (hence, McKinsey's own characterization of the approach as 'syntactical').

The final strand in the early development of modal semantics was the creation of possible worlds semantics. In some sense the most venerable approach, with (loose) roots going at least as far back as Leibniz, it was arguably first championed in the contemporary period by Carnap (1946, 1947/1956). The core concept in Carnap's semantics is that of a *state-description*, a set of sentences containing, for each atomic proposition, either that proposition or its negation (Carnap, 1947/1956, p. 9). He is quite explicit in connecting these state-descriptions to possible worlds:

---

[4] A *normal extension* of a modal logic **L** is a system containing **L** closed under substitution, modus ponens, and necessitation; the concept was deployed by McKinsey and Tarski (1948, p. 7) in connection with extensions of **S4**. Scroggs (1951) actually proves the stronger result that all proper *quasi-normal extensions* of **S5** (i.e., extensions of **S5** closed under modus ponens and substitution) are tabular, but, as he also shows, every quasi-normal extension of **S5** is normal.

[5] However, it may be remarked that Kripke (1963b, p. 69, n. 2) regarded Jónnson and Tarski (1951) as "the most surprising anticipation of the present [modal model] theory". For relevant discussion, consult Blackburn et al. (2001, pp. 40–41), Copeland (2002, Sect. 6, pp. 105–106), Goldblatt (2003, Sect. 3.4, pp. 324–325), and Burgess (2011, Sect. 7, pp. 128–129), as well as Lemmon (1966a, b) and the reviews thereof by Kripke (1967a, b). For a textbook treatment of the connections between Kripke semantics and the earlier algebraic approach, see Blackburn et al. (2001, Chap. 5).

[6] McKinsey (1945, p. 83, n. 3) states that his "syntactical definition of possibility is based on certain ideas" from Carnap (1937). However, it would be fair to say that the latter anticipates the former only rather loosely.

[7] See also Copeland (2002, Sects. 4, 12, 18) and, for some very recent developments, Bacon and Fine (in press). Copeland (2002, p. 120) indicates that Smiley's original work on the subject was done independently of McKinsey's. However, in a letter from January 31, 1959 to Kripke, Smiley states that his "main concern there [in his thesis] was to prove results essentially inspired by McKinsey's paper in JSL vol. 10, on interpretations of modal logic in terms of families of transformations." We have not been able to consult Smiley's dissertation.

> Thus the state-descriptions represent Leibniz' possible worlds or Wittgenstein's possible states of affairs. (Carnap, 1947/1956, p. 9)

Using $N$ as the sign for logical necessity, Carnap proposes the following convention: $N\varphi$ is true if and only if $\varphi$ is L-true, that is, if and only if $\varphi$ holds in every state-description (Carnap, 1947/1956, pp. 10, 174).

Carnap's approach can justifiably be described as a possible worlds semantics, but it lacks two key features that made Kripke's approach so versatile and powerful. First, Carnap has no notion of an accessibility relation, so his approach is effectively restricted to **S5**. Second, he has no notion of a frame (or *model structure*, in Kripke's preferred idiom)—one can only evaluate a modal expression against the set of absolutely all state-descriptions.[8]

Between Carnap and Kripke, a number of logicians, several working more-or-less independently, would develop the possible worlds semantic approach further. The history of these developments has been dealt with capably by other authors, including Copeland (2002, 2006), Goldblatt (2003), and Ballarin (2022), and so we will not dwell long on it here. The figures Kripke himself picks out as having anticipated his work in various ways include, among others, Bayart (1959), Hintikka,[9] and Kanger (1957),[10] with Hintikka being the one he refers to most frequently; consult Kripke (1959b, 1963a, p. 83, n. 1, 1963b, p. 69, n. 2). Prior also bears mentioning, since Kripke carried on an important correspondence with him on modal and tense logic in 1958 (see Sect. 2.4).

## 2.3 Charting Kripke's Path to Modal Logic

Before discussing how Kripke came to modal logic, some words are in order about how he came to logic. Although his own parents had already encouraged his interests in philosophy and mathematics through independent reading, it seems to have been a Mrs. Croft, an 8th grade mathematics teacher at Dundee Elementary School in

[8] The precise technical details of Carnap's account are beyond the scope of this piece. The interested reader should consult the technical sections of Carnap (1946, 1947/1956) as well as Copeland (2002, Sect. 5, pp. 104–105) and Goldblatt (2003, Sect. 4.3, pp. 329–330).

[9] It is not entirely clear in Kripke's earliest writing on modal logic what work of Hintikka's he has in mind in citing him as an antecedent; Kripke (1963a, p. 83, n. 1) cites Hintikka (1961) specifically, but Hintikka (1957) had introduced a sort of binary accessibility relation in considerably earlier work on quantified deontic logic.

[10] Kripke wrote to Kanger on January 24, 1958 to obtain a copy of Kanger (1957), which he received. As the letter makes apparent, however, Kripke's interest was not in the specifically modal contents of that work, and he does not appear to have been significantly influenced by its treatment of modal logic; on this, see also Copeland (2002, Sect. 20, pp. 130–131). Both Hintikka and Kanger are picked out as important antecedents by Kaplan (1966, p. 122), although he exaggerates the extent of Kanger's anticipation; see Copeland (2002, Sect. 13, pp. 121–123) and Goldblatt (2003, Sect. 4.5, pp. 334–335).

Omaha, Nebraska in the early 1950s, who first introduced him to logic and the foundations of mathematics.

Kripke became Mrs. Croft's student in the academic year 1952–1953. Recognizing that he was already beyond the established curriculum, she brought him books to read on his own while she taught the rest of the class. Among these books was *Mathematics and the Imagination* (Kasner & Newman, 1940).[11] The book introduces set theory and Cantor's diagonal argument, as well as paradoxes such as those of Banach-Tarski and Russell (Kasner & Newman, 1940, pp. 52ff., 205ff., 213ff.). It even alludes (not necessarily very precisely) to the foundational crisis in mathematics, gestures at what the views of the "Logistic", Intuitionist, and Formalist schools were, and briefly describes logical positivism (Kasner & Newman, 1940, pp. 61ff., 218ff., 359ff.). Already being interested in both philosophy and mathematics, Kripke was motivated to learn more about the 'mathematical philosophy' touched on in the book and, with it, formal logic.

Kripke was able to obtain copies of Quine (1951), Kleene (1952), and Rosser (1953) in fairly short order (cutting edge books, at the time).[12] He was later able to gain a deeper knowledge of the subject by taking trips to the University of Nebraska at Lincoln, which had a subscription to the *Journal of Symbolic Logic*. By the time he was 15 (1955/56), he had a subscription of his own.

The very first issue of the *Journal of Symbolic Logic* he received included "Modality and Quantification in **S5**" (Prior, 1956). Prior's note is devoted to showing that the Barcan formula, in the form $\Diamond(\exists x)\varphi(x) \rightarrow \exists x\Diamond\varphi(x)$, is already a theorem of **S5** extended with the usual quantificational axioms and so requires no separate postulation.[13] This brief note, written in Polish notation and devoid of any semantics, was apparently Kripke's first encounter with modern formal modal logic. He would go on to cite it in his first published article (Kripke, 1959a).[14]

Kripke began to think seriously about **S5** around that time and also began to read more widely in modal logic, becoming familiar with the other Lewis systems—probably, he recalled, by reading Lewis and Langford (1932/1959) among others. He devised a truth-table approach to **S5** utilizing tables with rows omitted, which

[11] Kripke recalled that Quine said to him years later, "That awful book—but maybe what we think of books like that should be more favorable if it led young people like you to get interested in the subject." It is perhaps somewhat remarkable that such a book should have discussed issues in logic at all; the subject was evidently a special interest of Newman's, who would later go on to coauthor Nagel and Newman (1958).

[12] Kripke recalled that he received Quine (1951) as a Bar Mitzvah present, so he must have had it no later than November 1953.

[13] Barcan (1946) had previously examined the formula in connection with **S2** as had Carnap (1946, p. 54) in connection with **MFC** (= quantified **S5**). However, Prior (1956) does not show any awareness of Carnap's work on this subject; see also Anderson (1957).

[14] He would return to it in connection with the problem of how to axiomatize quantified **S5** without the Barcan or converse Barcan formulae (Kripke, 1963a, pp. 88–90).

would later appear, perhaps with some refinements, in Kripke (1959a, pp. 11–12).[15] Other truth-table approaches to **S5** (and different modal logics or fragments thereof) had been studied around the same time or earlier, notably in conjunction with normal forms by von Wright (1951) in his *Essay on Modal Logic*.[16] Kripke recalled expressing considerable reservations about the originality of his work to Curry later, who assured him that, to the best of his knowledge, his approach was original.

In any case, Kripke was interested in showing his work on **S5** to an expert. In Berry (1956, p. 224), a meeting of the Association for Symbolic Logic to be held at the University of Rochester in Rochester, New York on December 27, 1956 was announced.[17] Kripke resolved to attend the meeting and show his work on **S5** to some of the logicians there.

Kripke's father, Rabbi Myer Kripke, arranged for him to stay with Rabbi Abraham Karp for the conference. Kripke's mother, Dorothy Kripke (a notable author of children's books), recalled:

> We stayed up all night before Saul was to leave for the convention […] We had two friends over to help us type the paper. Saul went to bed because we wanted him to be fresh, but we had to keep waking him up. We couldn't proofread the paper without him. None of us understood a word of it. (Branch, 1977)

A number of prominent logicians whom Kripke had been reading or with whom he would later correspond were at the conference in Rochester. Gál (1957) lists, among others, Anderson, Curry, Kleene, Kreisel, Myhill, and Putnam. Kripke recalled being too shy to approach anyone. Rabbi Karp suggested that he seek out Burton Dreben—notorious for being more Quinean than Quine and so hardly an ideal figure to speak to about modal logic—but he was not present.

Finally, Kripke suggested meeting Haskell Curry, which Rabbi Karp arranged. Kripke had by this point read at least some of Curry (1950), the fifth chapter of which deals with modal logic proof-theoretically.[18] Although Kripke was anxious about the meeting, Curry was evidently impressed with his work on **S5** and the meeting went well. The two would go on to have a fruitful correspondence and Kripke would later recall that Curry was the closest thing to a mentor he ever had. The thanks he gives to Curry, in particular, in his first published paper is worth quoting:[19]

---

[15] As he says there, he worked on the truth-table approach extensively before becoming acquainted with Beth's tableaux, which he ultimately used in obtaining his much better known completeness results (Kripke, 1959a, p. 12, n. 4).

[16] Although the informal character of von Wright (1951) makes it somewhat difficult to evaluate the exact connection between it and Kripke's approach, it does at least also involve row deletions. For an exposition, as well as corrections to various errata, see especially Anderson (1953). Anderson (1954) gives "improved" decision procedures for **M** (= **T**) and **S4** along the same lines.

[17] Assuming Kripke heard about this meeting from reading Berry (1956)—he certainly heard about it from reading the *Journal of Symbolic Logic*—and had already formulated his results on **S5** at that time, this would imply that Kripke had gone from having first encountered modal logic to obtaining his first completeness results in the subject in the span of about three months.

[18] The chapter's treatment of the subject is, however, technically defective; see Curry (1952).

[19] The referee referred to in the quote is A. N. Prior. Kripke learned this from Prior himself when they met in Helsinki in 1962.

My thanks to the referee and to Professor H. B. Curry for their helpful comments on this paper and their careful reading of it. I must express an added debt of gratitude to Curry; without his constant encouragement of my research, publication of these results might have been delayed for years. (Kripke, 1959a, p. 1, n. 1)

## 2.4 Kripke as Correspondent

Meeting Curry opened the doors for Kripke, who began to correspond not only with him but with numerous other logicians as well (often at Curry's direction). We will not attempt to treat Kripke's entire correspondence on modal logic here, but merely indicate how his thinking rapidly evolved even as his interests expanded in the critical period between meeting Curry and leading up to the publication of his first paper. We will also discuss some important later correspondence that indicates Kripke's enduring interest in the subject.

Kripke's early correspondence with Curry focused on proof-theoretic topics concerning the latter's work on the system **LD** as well as **S5**. (Kripke would be one of forty winners, or finalists, in the 1958 Westinghouse Science Talent Search with a paper submission on a system **LE** inspired by Curry's work. In Washington, D.C., for the competition, Kripke would meet Curry again, as well as Richard Nixon, then Vice President.) The correspondence makes it clear that it was Curry who put Kripke in contact with Beth, and so ultimately with Beth's work on tableaux.[20] It was also Curry who advised him to reach out to Kanger to obtain the latter's *Provability in Logic* (see Footnote 10).

Kripke initiated a correspondence with Alan Ross Anderson on deontic and modal logic not later than August 1957. Since Anderson's talk at the Rochester conference the preceding year was on deontic logic (Gál, 1957, p. 105), we may speculate that Kripke's thinking on the topic was stimulated from his having attended that talk. Anderson remarks in a letter from January 6, 1958 that he now regards the best deontic logic as being based on the system **E** of entailment. This was perhaps Kripke's first introduction to relevance logic. Kripke would shortly thereafter announce the positive solution to the decision problem for the pure implicational fragment of **E**—in Kripke (1959c), called **I**—his proof having been sketched in a letter to Nuel Belnap from September 21, 1959.[21]

Kripke carried on an important correspondence with Prior in the Fall of 1958.[22] Kripke had been reading Prior's recent book, *Time and Modality* (Prior, 1957), and noted an error in his treatment of **S4**. Roughly, Prior's idea was to represent a propo-

[20] It is not entirely clear from the extant correspondence what works Kripke received from Beth. In Kripke (1959a), he cites Beth (1955).

[21] Kripke would also refute Belnap's conjecture that the implicational fragment of **E** is the intersection of the implicational fragments of **R** and **S4**, put to him in a letter dated September 28, 1959. Kripke's counterexample would later be reproduced in Anderson and Belnap (1975, Sect. 8.10, p. 94). A related conjecture is refuted by Bimbó and Dunn (in press) in this volume.

[22] These letters were published in Ploug and Øhrstrøm (2012, pp. 372–379).

sition $p$'s truth value as a (ultimately, infinite) sequence of the values True and False (so, $p$'s being true now would correspond to the first item in the sequence being True, and so on); then $\Diamond p$ takes True at a point in the sequence just when $p$ takes True at that point or some later point in the sequence (Prior, 1957, pp. 21–23). The infinite matrix based on this idea is indeed quite natural,[23] but it does not, contrary to Prior's assertion, characterize **S4**. As Kripke points out, Prior's scheme validates $\Box\Diamond p \vee \Box\Diamond\neg p$, but this is clearly not a theorem of **S4** (Ploug & Øhrstrøm, 2012, p. 373).

It is, however, what Kripke then goes on to do that is most remarkable about his letter from September 3, 1958. He proposes a modified infinite "characteristic matrix" for **S4** based on a tree, with the root representing the present, its immediate successors the possibilities for the next moment, and so on (Ploug & Øhrstrøm, 2012, pp. 373–374). Kripke asserts that one can give denumerable tree matrices which are characteristic for **S4** as well as **T** and intuitionistic logic in a similar fashion (Ploug & Øhrstrøm, 2012, p. 374). These remarks prefigure the "tree" completeness results Kripke (1963b, 1965b) would publish for all of these logics only years later.[24]

In a subsequent letter to Prior from October 13, 1958, Kripke remarks that he has also found matrices for **E2** and **S2** which permit endpoints, "sorts of Armageddons after which time stops" (Ploug & Øhrstrøm, 2012, p. 376). At such endpoints, $\Box$-formulae are always false. Thus, already in 1958, Kripke had also obtained the central insight that underlies his semantics for non-normal modal logics, only to be published later in Kripke (1965a).

We conclude this section by discussing some considerably later correspondence of Kripke's germane to work of his appearing in this volume for the first time. The first item, reproduced in this volume as Kripke (in press-b), is a letter from Saul Kripke to David Lewis, dated August 11, 1969, concerning the latter's article "Counterpart Theory and Quantified Modal Logic" (Lewis, 1968). Lewis (1968) articulates an extensional first-order theory into which quantified modal logic is supposed to be translatable and which features, as its major conceptual innovation, a counterpart relation.[25] The innovation is philosophically driven: in response to Quinean qualms about unactualized possibilia and their identity conditions, Lewis (1968, p. 114) maintains that nothing exists in multiple different worlds after all—rather, things have counterparts, similar things which may stand in for them in other worlds.

---

[23] "If you remember that $Mp$ [$\Diamond p$] here means 'It is or will be the case that $p$', and suppose that in these sequences each place represents a time future to the one to the left of it, the rationale of this method of determination will be obvious" (Prior, 1957, p. 23). Incidentally, such a definition of possibility was explicitly endorsed by Diodorus (Boethius, *On Aristotle's De int.* 234, 22–26), a fact obviously known to Prior; see also his classic Prior (1955).

[24] In connection with the Prior correspondence, see especially Kripke (1963b, Sect. 5.2, pp. 92–94). Kripke recalled that he had already completed a draft of Kripke (1959a) before reading Prior (1957).

[25] The idea of modal counterparts (or a counterpart relation) occurs fuzzily as early as Leibniz (1885/1961, Sect. 414, pp. 362–363), from whom Geach evidently took inspiration (Copeland, 2002, p. 133, n. 21). Lewis (1968, p. 113, n. *) notes Geach's anticipation of his work in a letter Geach sent to Prior in 1964.

The first part of Kripke (in press-b) gestures at philosophical criticisms of counterpart theory he would make more fully (and famously) in *Naming and Necessity* (Kripke, 1972/1980). More relevant to the subject matter of this volume are Kripke's technical criticisms of Lewis's proposal. Kripke (in press-b) points out, among other things, that if individuals can have multiple counterparts in other worlds, Lewis's scheme will invalidate even $\forall y(\forall x\varphi(x) \rightarrow \varphi(y))$.[26]

Second, in connection with Kripke (in press-a), which appears for the first time in this volume,[27] Kripke reported technical results concerning modal logic with quantification over 'individual concepts' in a letter to Kaplan and Thomason dated March 12, 1976. Now, in standard quantified modal logic, individual variables and terms are assigned to *individuals*, that is, elements of a domain. Carnap (1947/1956, Sect. 40, p. 180) held that "individual variables in modal sentences [...] must be interpreted as referring, not to individuals, but to individual concepts." An *individual concept* is the intension of an individual expression or, more concretely, a function from worlds to individuals (Carnap, 1947/1956, pp. 41, 181).[28] In his letter to Kaplan and Thomason, Kripke reported that, in contrast to the unaxiomatizability of many standard modal logics with quantification over individual concepts, **S5** with individual concepts is axiomatizable; he then axiomatizes it and proves strong unaxiomatizability results for other standard modal logics.[29]

Although Kripke published little on modal logic after 1965 (and, as we have indicated, most of the results he published in the 1960s were in fact obtained in the 1950s), the correspondence just referred to shows that he continued to actively work on the subject well into the 1970s. It also indicates Kripke's enduring interest in both philosophical aspects of, and mainly technical problems in, modal logic.

---

[26] For some discussion of this letter and the problems Kripke (and others) have noted with Lewis's proposal, see Dorr et al. (2021, Sect. 10.2). See also Kripke (1972/1980, p. 45, n. 13).

[27] An extended abstract was previously published as Kripke (1992).

[28] The details are technical; consult Hughes and Cresswell (1996, Chap. 18) for a relatively accessible introduction on how to actually build models of such things.

[29] The history of unaxiomatizability results for quantified modal logic with (all) individual concepts is rather opaque. Briefly, the first mention of a pertinent result in the literature seems to be by Thomason (1969, p. 131), who attributes to Kaplan the result that the validities of **Q2** are not recursively enumerable; he further asserts that the result depends on work from Kaplan (1964). See also Thomason (1970, p. 75, n. 1), where he adds that Kaplan reported this result to him in correspondence from April 1967. However, Kaplan's argument was subsequently found to be erroneous, which precipitated Kripke's own investigations in the subject (Bacon, 1980, p. 194). By the late 1970s, Kamp (1977) was being informally circulated; this includes proofs of two unaxiomatizability results, one in quantified tense logic (attributed to Dana Scott) and the other in modal logic with quantification over individual concepts (attributed to Saul Kripke). These results are discussed in Garson (1984, Sect. 3). See also the editors' introduction to Kripke (in press-a).

## 2.5 Kripke's Papers

In this section we canvass Kripke's published output on modal logic to give the unfamiliar reader a very general sense of its character and content. We will, again, not treat everything Kripke wrote on the subject, but focus on the seminal papers whose results he obtained in the late 1950s as well as a selection of lesser known later work.

Kripke's first published article, Kripke (1959a), contains as its main results completeness theorems for propositional **S5**, **S5** with constant domain quantifiers, and **S5** with constant domain quantifiers and identity. The technique Kripke uses to prove completeness is based on the semantic tableaux of Beth (1955), which are distinct from, though naturally related to, the sort of tableaux one typically encounters in textbooks on the subject today. The article neither invented nor used the now standard Henkin-style canonical model technique for proving modal completeness theorems.[30]

The footer of Kripke (1959a, p. 1) notes that the article was received by the *Journal of Symbolic Logic* on August 25th, 1958. This is presumably the date that it was received in a revised form, since a letter from Curry to Kripke dated July 22, 1958 suggests that Kripke's article had already been accepted earlier—Kripke had reached out to Curry on July 18th, 1958 for advice on revisions and Curry prodded him not to change too much and to reserve an additional result for another paper.

Kripke (1959b), a short abstract, announces completeness theorems for most of the usual normal and non-normal modal logics—both propositional and quantified—that logicians up to that point had been interested in. It also announces simple decision procedures for several of the (propositional) systems and a semantics for intuitionistic logic based on that for **S4**. It further alludes to "certain metamathematical applications of modal logic" (Kripke, 1959b, p. 324).[31] The abstract was received by the *Journal of Symbolic Logic* on October 21st, 1959.

It would be several years before Kripke's next publication on the subject (Kripke, 1962), or before he published articles backing up the claims of his 1959 abstract (Kripke 1963a, b, 1965a, b). This delay is easily explained. In 1958, Kripke went to Harvard to study mathematics and, under Quine's influence, work on modal logic was discouraged there. Nevertheless, it should be emphasized, work on modal logic was by no means prohibited. Kripke (2017a) was originally written for a seminar

[30] The origin of this technique is also unclear. Kaplan (1966, pp. 121–122) briefly sketches the argument and states that it had been suggested to him by Dana Scott. Around the same time, Makinson (1966) published proofs of some completeness results in modal logic using the same technique. He further remarks that the results are based on a section of his 1965 D. Phil. thesis (Makinson, 1966, p. 379, n. 1).

[31] It is not entirely clear what Kripke envisioned here. Kripke (1963a, pp. 90–92) briefly considers provability interpretations of modal logic connected with Peano arithmetic. It is also noteworthy (and was later noted by Kripke) that the method of 'forcing', developed by Cohen (1963) to prove the independence of the Continuum Hypothesis, can be given an exposition in terms of Kripke's semantics (Kripke, 1965b, pp. 118–120).

given by Quine himself in 1961/1962, although it was only published much later. It deals most directly with Quine's critical views on modal logic as expressed in Quine (1953/1976).

As we have indicated, Kripke (1963b, 1965a) accomplishes much of the technical agenda that was advertised in Kripke's earlier abstract—the central results are again proved using tableaux. Kripke (1963a), on the other hand, is more of a conceptual piece. Among other things, it treats variable domain quantification, considers the problem of how to handle $\varphi(x)$ when the variable $x$ is assigned to an individual not in a given world's domain, and shows how to build models rejecting both the Barcan and converse Barcan formulae.

After 1965, Kripke published on modal logic only sparingly. He wrote several reviews of other logicians' work, including Kripke (1967a, b) on Lemmon (1966a, b) and Kripke (1983, 1985) on Fine (1978a, b, 1979, 1981). As reviews go, these pieces are all rather substantial and demonstrate Kripke's continued serious engagement with the field (and with issues involving modality and quantification, in particular). For example, Kripke (1967a) answers various problems left open by Lemmon (1966a, b) and questions how Lemmon's methods might be extended to quantified modal logic (along the way introducing a modal logic which poses no issues at the propositional level but does with the introduction of quantifiers). Besides these, we have already mentioned the abstract published as Kripke (1992) and the seminar paper published as Kripke (2017a), to which he also appended some afterthoughts in Kripke (2017b).

## 2.6 Concluding Remarks

Kripke did not invent possible worlds semantics for modal logic (broadly understood), let alone the concept of a 'possible world', but his singular innovations and insights plausibly warrant crediting him as the creator of the first truly general modal *model theory*.[32] That he was able to accomplish much of this as a teenager with relatively limited—though, as we saw, nevertheless quite important—interaction with the wider logic community makes that accomplishment all the more impressive. We conclude by remarking on a couple of themes that emerge from the narrative of Kripke's work on modal logic.

The first theme, fittingly, is that of *contingency*. What *if* Mrs. Croft had not given Kripke a book to read that discussed logic? What *if* Kripke had met Burton Dreben in Rochester instead of Haskell Curry? Counterfactuals are tricky, but likely the history of modal logic would have been very different. That Kripke accomplished what he did depended in no small part on his having or finding supportive figures who encouraged his interests and work—a fact he himself certainly recognized.

[32] See Burgess (2011, Sect. 7) for a similar assessment.

Our second theme is that of the tight integration of philosophical and technical thinking in Kripke's work. Perhaps this is most apparent in connection with his correspondence with Prior, where Kripke's trees have an obvious tense-logical interpretation in terms of branching time; it is tempting to suspect that thinking on that philosophical idea preceded the completeness insight, even if Kripke suggests it did not (Ploug & Øhrstrøm, 2012, pp. 373–374).

However, this theme is also exhibited in Kripke's long-term fascination with issues specific to *quantification* in modal logic. In the Quinean atmosphere of his youth, it was quantified modal logic, especially, that was considered the most philosophically dubious.[33] Kripke expended considerable effort in formally rehabilitating quantified modal logic—effort which would lay a sort of groundwork for his later epochal philosophical work.

**Acknowledgements** We would like to thank Mircea Dumitru, Melvin Fitting, Oliver Marshall, Gary Ostertag, Martin Pleitz, and Graham Priest for comments on previous drafts.

## References

Anderson, A. R. (1953). Review of Georg H. v. Wright, *An essay in modal logic* (Studies in logic and the foundations of mathematics, North-Holland publishing company, Amsterdam 1951, VII + 90 pp). *Journal of Symbolic Logic, 18*(2), 174–176.

Anderson, A. R. (1954). Improved decision procedures for Lewis's calculus S4 and von Wright's calculus M. *Journal of Symbolic Logic, 19*(3), 201–214.

Anderson, A. R. (1957). Review of A. N. Prior, 'Modality and Quantification in S5' (Journal of Symbolic Logic, 21 (1956), pp. 60–62). *Journal of Symbolic Logic, 22*(1), 91.

Anderson, A. R., & Belnap, N. D. (1975). *Entailment: The logic of relevance and necessity*. Princeton: Princeton University Press.

Bacon, A., & Fine, K. (in press). The logic of logical necessity. In Y. Weiss & R. Birman (Eds.), *Saul Kripke on modal logic*. Cham: Springer.

Bacon, J. (1980). Substance and first-order quantification over individual-concepts. *Journal of Symbolic Logic, 45*(2), 193–203.

Ballarin, R. (2022). Modern origins of modal logic. Retrieved 12 July 2023, from https://plato.stanford.edu/archives/win2022/entries/logic-modal-origins/

Barcan, R. C. (1946). A functional calculus of first order based on strict implication. *Journal of Symbolic Logic, 11*(1), 1–16.

Bayart, A. (1959). Quasi-adéquation de la logique modale du second ordre S5 et adéquation de la logique modale du premier ordre S5. *Logique et Analyse, 2*(6/7), 99–121.

Berry, G. D. W. (1956). Twentieth meeting of the Association for Symbolic Logic. *Journal of Symbolic Logic, 21*(2), 216–224.

Beth, E. W. (1955). Semantic entailment and formal derivability. *Mededelingen der Koninklijke Nederlandse Akademie van Wetenschappen, Afd Letterkunde, 18*(13), 309–342.

Bimbó, K., & Dunn, J. M. (in press). Entailment, mingle and binary accessibility. In Y. Weiss & R. Birman (Eds.), *Saul Kripke on modal logic*. Cham: Springer.

[33] After all, neither invidious "Aristotelian" essentialism nor the famous Slingshot Argument really seem to threaten the legitimacy of *propositional* modal logic much.

Blackburn, P., de Rijke, M., & Venema, Y. (2001). *Modal logic*. Cambridge: Cambridge University Press.
Branch, T. (1977). New frontiers in American philosophy. *New York Times* (August 14), 180.
Burgess, J. (2011). Kripke models. In A. Berger (Ed.), *Saul Kripke* (pp. 119–140). Cambridge: Cambridge University Press.
Carnap, R. (1937). *The logical syntax of language.* A. Smeaton (Trans., Ed.). London: Routledge & Kegan Paul Ltd.
Carnap, R. (1946). Modalities and quantification. *Journal of Symbolic Logic, 11*(2), 33–64.
Carnap, R. (1947/1956). *Meaning and necessity: A study in semantics and modal logic*. Chicago: Phoenix Books (University of Chicago Press).
Cohen, P. J. (1963). The independence of the continuum hypothesis. *Proceedings of the National Academy of Sciences, U.S.A*, *50*, 1143–1148.
Copeland, B. J. (2002). The genesis of possible worlds semantics. *Journal of Philosophical Logic, 31*(2), 99–137.
Copeland, B. J. (2006). Meredith, Prior, and the history of possible worlds semantics. *Synthese, 150*(3), 373–397.
Curry, H. B. (1950). *A theory of formal deducibility*. Notre Dame: University of Notre Dame.
Curry, H. B. (1952). The elimination theorem when modality is present. *Journal of Symbolic Logic, 17*(4), 249–265.
Dorr, C., Hawthorne, J., & Yli-Vakkuri, J. (2021). *The bounds of possibility: Puzzles of modal variation*. Oxford: Oxford University Press.
Drake, F. R. (1962). On McKinsey's syntactical characterizations of systems of modal logic. *Journal of Symbolic Logic, 27*(4), 400–406.
Dugundji, J. (1940). Note on a property of matrices for Lewis and Langford's calculus of propositions. *Journal of Symbolic Logic, 5*(4), 150–151.
Fine, K. (1978a). Model theory for modal logic part I-The *De Re/De Dicto* distinction. *Journal of Philosophical Logic, 7*(1), 125–156.
Fine, K. (1978b). Model theory for modal logic-part II: The elimination of *De Re* modality. *Journal of Philosophical Logic, 7*(1), 277–306.
Fine, K. (1979). Failures of the interpolation lemma in quantified modal logic. *Journal of Symbolic Logic, 44*(2), 201–206.
Fine, K. (1981). Model theory for modal logic-part III: Existence and predication. *Journal of Philosophical Logic, 10*(3), 293–307.
Gál, L. (1957). Twenty-first annual meeting of the Association for Symbolic Logic. *Journal of Symbolic Logic, 22*(1), 105–112.
Garson, J. W. (1984). Quantification in modal logic. In D. Gabbay & F. Guenthner (Eds.), *Extensions of classical logic* (1st ed., Vol. 2, pp. 249–307). Dordrecht: D. Reidel Publishing Company.
Gödel, K. (1933/1986). Eine interpretation des intuitionistischen aussagenkalküls. In S. Feferman, J. W. Dawson, Jr., S. C. Kleene, G. H. Moore, R. M. Solovay, & J. van Heijenoort (Eds.), *Kurt Gödel: Collected works* (Vol. 1, pp. 300–303). New York: Oxford University Press.
Goldblatt, R. (2003). Mathematical modal logic: A view of its evolution. *Journal of Applied Logic, 1*(5–6), 309–392.
Hintikka, J. (1957). Quantifiers in deontic logic. *Societas Scientiarum Fennica, Commentationes humanarum litterarum*, *23*(4).
Hintikka, J. (1961). Modality and quantification. *Theoria, 27*(3), 119–128.
Hughes, G. E., & Cresswell, M. J. (1996). *A new introduction to modal logic*. London: Routledge.
Jónnson, B., & Tarski, A. (1951). Boolean algebras with operators I. *American Journal of Mathematics, 73*(4), 891–939.
Kamp, H. (1977). Two related theorems by D. Scott and S. Kripke. (Mimeographed.)
Kanger, S. (1957). *Provability in logic*. Stockholm: Almqvist and Wiksell.
Kaplan, D. (1964). *Foundations of intensional logic* (Unpublished doctoral dissertation). Los Angeles: University of California.

Kaplan, D. (1966). Review of Saul A. Kripke, 'Semantical analysis of modal logic I. Normal modal propositional calculi' (Zeitschrift für mathematische Logik und Grundlagen der Mathematik, 9 (1963), pp. 67–96). *Journal of Symbolic Logic, 31*(1), 120–122.

Kasner, E., & Newman, J. (1940). *Mathematics and the imagination*. New York: Simon and Schuster.

Kleene, S. C. (1952). *Introduction to metamathematics*. New York: D. Van Nostrand Company.

Kneale, W., & Kneale, M. (1962). *The development of logic*. Oxford: Clarendon Press.

Kripke, S. A. (1959a). A completeness theorem in modal logic. *Journal of Symbolic Logic, 24*(1), 1–14.

Kripke, S. A. (1959b). Semantical analysis of modal logic (abstract). *Journal of Symbolic Logic, 24*(4), 323–324.

Kripke, S. A. (1959c). The problem of entailment (abstract). *Journal of Symbolic Logic*, *24*(4), 324.

Kripke, S. A. (1962). The undecidability of monadic modal quantification theory. *Zeitschrift für mathematische Logik und Grundlagen der Mathematik, 8*, 113–116.

Kripke, S. A. (1963a). Semantical considerations on modal logic. *Acta Philosophica Fennica, 16*, 83–94.

Kripke, S. A. (1963b). Semantical analysis of modal logic I. Normal modal propositional calculi. *Zeitschrift für mathematische Logik und Grundlagen der Mathematik, 9*, 67–96.

Kripke, S. A. (1965a). Semantical analysis of modal logic II. Non-normal modal propositional calculi. In J. W. Addison, L. Henkin & A. Tarski (Eds.), *The Theory of Models, Proceedings of the 1963 International Symposium at Berkeley* (pp. 206–220). Amsterdam: North Holland Publishing Company.

Kripke, S. A. (1965b). Semantical analysis of intuitionistic logic I. In J. N. Crossley & M. A. E. Dummett (Eds.), *Formal Systems and Recursive Functions, Proceedings of the Eighth Logic Colloquium at Oxford, July, 1963* (pp. 92–130). Amsterdam: North Holland Publishing Company.

Kripke, S. A. (1967a). Review of E. J. Lemmon, 'Algebraic semantics for modal logic II' (Journal of Symbolic Logic, 31 (1966), pp. 191–218). *Mathematical Reviews, 34*(5), 1022.

Kripke, S. A. (1967b). Review of E. J. Lemmon, 'Algebraic semantics for modal logic I'(Journal of Symbolic Logic, 31 (1966), pp. 46–65). *Mathematical Reviews*, *34*(5), 1021–1022.

Kripke, S. A. (1972/1980). *Naming and necessity*. Cambridge: Harvard University Press.

Kripke, S. A. (1983). Review of Kit Fine, 'Failures of the interpolation lemma in quantified modal logic' (Journal of Symbolic Logic, 44 (1979), pp. 201–206). *Journal of Symbolic Logic, 48*(2), 486–488.

Kripke, S. A. (1985). Review of Kit Fine, 'Model theory for modal logic. Parts I–III' (Journal of Philosophical Logic, 7 (1978), pp. 125–156, 277–306; and 10 (1981), pp. 293–307). *Journal of Symbolic Logic, 50*(4), 1083–1093.

Kripke, S. A. (1992). Individual concepts: Their logic, philosophy, and some of their uses (abstract). *Proceedings and Addresses of the American Philosophical Association, 66*(2), 70–73.

Kripke, S. A. (2017a). Quantified modality and essentialism. *Noûs, 51*(2), 221–234.

Kripke, S. A. (2017b). Quantified modal logic and Quine's critique: Some further observations. *Noûs, 51*(2), 235–237.

Kripke, S. A. (in press-a). Individual concepts: Their logic, philosophy, and some of their uses. In Y. Weiss & R. Birman (Eds.), *Saul Kripke on modal logic.* Cham: Springer.

Kripke, S. A. (in press-b). A letter from Kripke to Lewis. In Y. Weiss & R. Birman (Eds.), *Saul Kripke on modal logic.* Cham: Springer.

Leibniz, G. W. (1885/1961). Essais de Théodicée. In C. I. Gerhardt (Ed.), *Die philosophischen schriften von Gottfried Wilhelm Leibniz* (Vol. 6). Hildesheim: Georg Olms.

Lemmon, E. J. (1957). New foundations for Lewis modal systems. *Journal of Symbolic Logic, 22*(2), 176–186.

Lemmon, E. J. (1966a). Algebraic semantics for modal logics I. *Journal of Symbolic Logic, 31*(1), 46–65.

Lemmon, E. J. (1966b). Algebraic semantics for modal logics II. *Journal of Symbolic Logic, 31*(2), 191–218.

Lewis, C. I., & Langford, C. H. (1932/1959). *Symbolic logic* (2nd ed.). New York: Dover.

Lewis, D. (1968). Counterpart theory and quantified modal logic. *Journal of Philosophy, 65*(5), 113–126.
Łukasiewicz, J. (1920/1970). On three-valued logic. In L. Borkowski (Ed.), *Jan Łukasiewicz: Selected works* (pp. 87–88). Amsterdam: North Holland Publishing Company.
Łukasiewicz, J. (1930/1970). Philosophical remarks on many-valued systems of propositional logic. In L. Borkowski (Ed.), *Jan Łukasiewicz: Selected works* (pp. 153–178). Amsterdam: North Holland Publishing Company.
Makinson, D. (1966). On some completeness theorems in modal logic. *Zeitschrift für mathematische Logik und Grundlagen der Mathematik, 12*, 379–384.
Malink, M. (2013). *Aristotle's modal syllogistic*. Cambridge: Harvard University Press.
McKinsey, J. C. C. (1941). A solution to the decision problem for the Lewis systems S2 and S4, with an application to topology. *Journal of Symbolic Logic, 6*(4), 117–134.
McKinsey, J. C. C. (1945). On the syntactical construction of systems of modal logic. *Journal of Symbolic Logic, 10*(3), 83–94.
McKinsey, J. C. C., & Tarski, A. (1948). Some theorems about the sentential calculi of Lewis and Heyting. *Journal of Symbolic Logic, 13*(1), 1–15.
Nagel, E., & Newman, J. (1958). *Gödel's proof*. New York: New York University Press.
Parry, W. T. (1934). The postulates for "Strict Implication". *Mind, 43*(169), 78–80.
Ploug, T., & Øhrstrøm, P. (2012). Branching time, indeterminism and tense logic: Unveiling the Prior-Kripke letters. *Synthese, 188*(3), 367–379.
Prior, A. N. (1955). Diodoran modalities. *Philosophical Quarterly, 5*(20), 205–213.
Prior, A. N. (1956). Modality and quantification in S5. *Journal of Symbolic Logic, 21*(1), 60–62.
Prior, A. N. (1957). *Time and modality*. Oxford: Clarendon Press.
Quine, W. V. (1951). *Mathematical logic*. Cambridge: Harvard University Press.
Quine, W. V. (1953/1976). Three grades of modal involvement. In W. V. Quine (Ed.), *The ways of paradox and other essays* (pp. 158–176). Cambridge: Harvard University Press.
Rosser, J. B. (1953). *Logic for mathematicians*. New York: McGraw-Hill.
Scroggs, S. J. (1951). Extensions of the Lewis system S5. *Journal of Symbolic Logic, 16*(2), 112–120.
Smiley, T. J. (1956). *Natural systems of logic* (Unpublished doctoral dissertation). Cambridge: Clare College, University of Cambridge.
Thomason, R. H. (1969). Modal logic and metaphysics. In K. Lambert (Ed.), *The logical way of doing things* (pp. 119–146). New Haven: Yale University Press.
Thomason, R. H. (1970). Some completeness results for modal predicate calculi. In K. Lambert (Ed.), *Philosophical problems in logic: Some recent developments* (pp. 56–76). Dordrecht: D. Reidel Publishing Company.
von Wright, G. H. (1951). *An essay in modal logic*. Amsterdam: North Holland Publishing Company.

# Chapter 3
# Beyond Knowledge of the Model

**Sergei Artemov**

**Abstract** The principle motivation of this work is to display the sense in which the epistemic reading of a Kripke model tacitly requires *common knowledge of the model, CKM*. This requirement significantly restricts the amount of epistemic situations we are able to consider. We explore possible worlds epistemic models in a general setting without *CKM* assumptions and show that such models can be identified with observable substructures of Kripke models. We argue that such *observable models* offer a new level of generality and conceptual clarity in epistemic modeling. Similar analysis applies to intuitionistic models.

**Keywords** Canonical model · Common knowledge · Epistemology · Knowledge · Kripke model · Intuitionistic model

## 3.1 Motivations

We now discuss a foundational problem with the standard use of Kripke models in epistemology.[1] Here is a quote from Williamson (2015, pp. 131–132):

> *Informally, we interpret W as a set of mutually exclusive, jointly exhaustive worlds or states, . . . R is a relation of epistemic accessibility: a world w has R to a world x if and only if . . . whatever the agent knows in w is true in x. We define a function* **K** *from propositions to propositions by the following equation for all propositions p*:
>
> $$\mathbf{K}p = \{w \in W : \forall x \in W, wRx \Rightarrow x \in p\}. \tag{3.1}$$
>
> *In other words,* $\mathbf{K}p$ *is true at a world if and only if p is true at every world epistemically accessible from that one. Informally,* $\mathbf{K}p$ *is interpreted as the proposition that the agent knows p.*

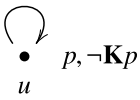
[1] For preliminary versions of this paper, cf. Artemov (2016, 2020). Since then, the text has been substantially revised and extended.

S. Artemov (✉)
The City University of New York, Graduate Center, 365 5th Avenue, New York, NY 10016, USA
e-mail: sartemov@gc.cuny.edu

© The Author(s), under exclusive license to Springer Nature Switzerland AG 2024

Y. Weiss and R. Birman (eds.), *Saul Kripke on Modal Logic*, Outstanding Contributions to Logic 30, https://doi.org/10.1007/978-3-031-57635-5_3

**Fig. 3.1** Structure $\mathcal{M}_1$




In notation of Williamson (2015), $R$ is defined here via knowledge at the states in $W$ as

$$R(w) = \{x \in W : \forall p\ (\text{“}p \textit{ is known at } w\text{”} \Rightarrow p \in x)\}. \tag{3.2}$$

Condition (3.2) yields

$$\text{“}p \textit{ is known at } w\text{”} \quad \Rightarrow \quad \textit{for all } x \in R(w),\ p \in x. \tag{3.3}$$

However, (3.2) does not guarantee the converse:

$$(\textit{for all } x \in R(w),\ p \in x) \quad \Rightarrow \quad \text{“}p \textit{ is known at } w,\text{”} \tag{3.4}$$

which is, nevertheless, built into definition (3.1) presuming $\mathbf{K}p$ is meant to formalize *“$p$ is known.”* So, (3.1) tacitly assumes both (3.3) and (3.4) whereas the latter does not follow from the definition of epistemic accessibility (3.2) without additional explicit (meta) assumptions.

As a result, Eqs. (3.1) and (3.2) do not match mathematically if we want to read “$w \in \mathbf{K}p$” as *“$p$ is known at $w$.”* Given knowledge assertions *“$p$ is known at $w$,”* we indeed can find accessibility relation $R$ by (3.2), and then determine the knowledge modality $\mathbf{K}$ by (3.1). The problem is that this $\mathbf{K}p$ can be different from the original knowledge assertion *“$p$ is known.”*

**Example 3.1.1** (*technical*) Consider S5 with a single propositional letter $p$ and a structure $\mathcal{M}_1$ consisting of one state $u$ generated by $\Gamma = \{p, \neg\mathbf{K}p\}$ (Fig. 3.1).[2]

The induced accessibility relation $R$, (3.2), is $uRu$, the truth at $u$ is membership in $u$:

$$u \models F \quad \Leftrightarrow \quad F \in u.$$

In particular, $\mathbf{K}p$ is false at $u$. On the other hand, by (3.1), $\mathbf{K}p$ ought to be true at $u$. So (3.2) and (3.1) are not compatible in $\mathcal{M}_1$.

While meta-knowledge of the model by the agent is assumed in epistemic reasoning, this can conflict with the traditional way of using Kripke models. The following example was discussed in Liu (2019) as a puzzle with some story attached to it. We consider only the formal side of the puzzle (Fig. 3.2).

**Example 3.1.2** (*consistency concerns*) Assume the Kripke model $\mathcal{M}_2$ for an S4-knower. The agent at $u$ considers both worlds $u$ and $v$ possible. Since the knower

[2] State $u$ is constructive: one can check that $\Gamma = \{p, \neg\mathbf{K}p\}$ is a complete set of formulas, i.e., for each $F$, either $\Gamma$ proves $F$ or $\Gamma$ proves $\neg F$, and $u$ is the set of formulas derivable from $\Gamma$, cf. also Sect. 3.8.

**Fig. 3.2** Model $\mathcal{M}_2$

knows the model, she realizes that at $v$ she would not consider $u$ possible and concludes that $v$ is not possible at $u$. This yields that the agent knows the real world $u$, $p$ at $u$, which is not what is suggested by the model. So, knowledge of the model is inconsistent with the epistemic reading of the model.

We postpone the solution of this puzzle until Sect. 3.10.

Kripke models provide a natural and convenient method for specifying worlds as maximal consistent sets of formulas. The downside of Kripke specification is that some nodes in a "natural model" of a given scenario might not be epistemically possible. In particular, in order to model ignorance of a fact $F$ by Kripke, one has to commit to a hypothetic accessible world at which $F$ is false. Such a world may be epistemically unattainable.

**Example 3.1.3** (*conceptual*) An agent respects mathematics but does not know enough of it to know Fermat's Little Theorem *FLT* (provable in Peano Arithmetic PA). In a Kripke model, we have to have a world deemed possible by the agent at which *FLT* is false. In such a world, the agent has to learn enough of PA to make sense of *FLT* and then to justify that *FLT* could be false, which is impossible without falsifying PA. This is, indeed, an impossible requirement for even the most brilliant of agents.

How do we get out of this predicament? We suggest by making use of epistemic models which naturally can allow a proposition to be true at each possible world while remaining unknown.

## 3.2 Epistemic Structures

On the formal side, our default is the propositional $n$-agent epistemic logic $\mathsf{S5}^n$, cf. Fagin et al. (1995), though all the main findings and suggestions apply to other modal logics as well. Similar considerations apply to other classes of models, e.g., Aumann structures (Aumann, 1995).

The language of $n$-agent modal logic is the propositional language with Boolean connectives and modalities $\mathbf{K}_1, \mathbf{K}_2, \ldots, \mathbf{K}_n$ which we will read as knowledge modalities for agents $1, 2, \ldots, n$. Postulates of $\mathsf{S5}^n$ are axioms and rules of classical propositional logic together with specific modal principles, for each $i = 1, 2, \ldots, n$,

1. $\mathbf{K}_i(P \to Q) \to (\mathbf{K}_i P \to \mathbf{K}_i Q)$ (distributivity)
2. $\mathbf{K}_i P \to P$ (factivity)
3. $\mathbf{K}_i P \to \mathbf{K}_i \mathbf{K}_i P$ (positive introspection)

4. $\neg \mathbf{K}_i P \to \mathbf{K}_i \neg \mathbf{K}_i P$ (negative introspection)
5. Necessitation Rule: $\vdash F \ \Rightarrow \ \vdash \mathbf{K}_i F$.

For $\mathsf{S4}^n$, remove the negative introspection axiom.

**Definition 3.2.1** For a given set of formulas $\Gamma$ (here called "hypotheses" or "assumptions") we consider *derivations from* $\Gamma$: assume all $\mathsf{S5}^n$-theorems together with $\Gamma$ and use classical propositional reasoning (rule *Modus Ponens*). The notation

$$\Gamma \vdash A$$

represents '$A$ is derivable from $\Gamma$.'

The Necessitation rule is not valid in a general setting for derivations from assumptions: for some $\Gamma$, $\Gamma \vdash F$ does not yield $\Gamma \vdash \mathbf{K}_i F$. Indeed, for any proposition $P$, $P \vdash P$, but, generally speaking $P \nvdash \mathbf{K}_i P$.

Therefore, when speaking about derivations from hypotheses, we do not postulate Necessitation. For some "good" $\Gamma$'s, Necessitation is a valid rule though, cf. Sect. 3.8.

Logic $\mathsf{S5}^n$ is sound and complete with respect to Kripke models with equivalence relations $R_1, R_2, \ldots, R_n$ as accessibility relations. For $\mathsf{S4}^n$, the accessibility relations $R_1, R_2, \ldots, R_n$ are reflexive and transitive.

Informally, by a *global state* of a multi-agent system we understand a complete description of epistemic states of agents along with the state of nature, represented as a set of propositions in an appropriate epistemic language.

Formally, *global state* (also *state* or *world*) is a maximal consistent (over a given logic base) set of formulas. The truth assignment "$\models$" to formulas at a world $w$ is defined as

$$w \models F \quad \textit{iff} \quad F \in w.$$

We begin by defining epistemic structures for the single agent case, as outlined in the quote in Sect. 3.1.

**Definition 3.2.2** **An epistemic structure** is a triple $(W, R, \models)$ in which

- $W$ is a nonempty set of elements which are called states (possible worlds);
- $\models$ is a complete truth relation at each world respecting the base logic: for each $u \in W$, $\{F \mid u \models F\}$ is a maximal consistent set;
- $R$ is a binary relation on $W$ s.t.

$$w \models \mathbf{K}F \quad \Rightarrow \quad R(w) \models F.$$

**A**[3] **Kripke model** $(W, R, \Vdash)$ **associated with** $(W, R, \models)$ is a Kripke model with the frame $(W, R)$ and atomic forcing relation "$\Vdash$":

$$u \Vdash p \quad \textit{iff} \quad u \models p.$$

[3] This is, of course, a version of (3.3) from Sect. 3.1.

Truth assignments in $\models$ and in $\Vdash$ coincide for the modal-free formulas, but can differ for general modal formulas as we can see in Example 3.1.1.

**Definition 3.2.3** An epistemic structure $(W, R, \models)$ is **fully observable** if for each $w$ and $F$,

$$R(w) \models F \quad \Rightarrow \quad w \models \mathbf{K}F.$$

In an epistemic structure, once $F$ is known at $u$ (i.e., $u \models \mathbf{K}F$), then $F$ holds everywhere in $R(u)$. For fully observable epistemic structures, the converse also holds: each $F$ which is true at all states accessible from $u$ ought to be known at $u$.

**Example 3.2.4** Epistemic structure $\mathcal{M}_1$ from Example 3.1.1 is not fully observable.

**Proposition 3.2.5** *An epistemic structure $\mathcal{M}$ coincides with its associated Kripke model iff $\mathcal{M}$ is fully observable.*

***Proof*** Any Kripke model is a fully observable.

Let an epistemic structure

$$(W, R, \models)$$

be fully observable and $(W, R, \Vdash)$ be the associated Kripke model. By induction on formula $F$ we check that for each $u \in W$,

$$u \models F \quad \Leftrightarrow \quad u \Vdash F.$$

The claim is secured by definitions for atomic $F$'s and obvious for the Boolean steps. Let $F$ be $\mathbf{K}X$. If $u \models \mathbf{K}X$ then $R(w) \models X$. By IH, $R(w) \Vdash X$, hence $u \Vdash \mathbf{K}X$. If $u \Vdash \mathbf{K}X$ then $R(w) \Vdash X$. By IH, $R(w) \models X$, hence, by full observability, $u \models \mathbf{K}X$. □

So, Kripke models are exactly fully observable epistemic structures.

Here is an informal sufficient condition under which an epistemic structure is a Kripke model (a fully observable epistemic structure).[4] A Kripke model is an epistemic structure which is known to the agent, along with the fact that it provides a comprehensive description of a given epistemic situation.

Once $F$ holds everywhere in $R(u)$, the agent knows this and, knowing $R$, can conclude that $F$ holds at all states epistemically possible in $u$, thus coming to the justified (by virtue of this argument) knowledge of $F$. So, in Kripke models, knowledge of $F$ at $u$, given that $F$ holds in $R(u)$, does not come from nowhere. The justification for knowledge

$$u \Vdash \mathbf{K}F$$

is merely assumed knowledge of the model itself relativized to a specific state $u$.

[4] It appears that a natural formalization of this condition leads us beyond the current level of propositional modal logic. It would be interesting to see it done though.

**Definition 3.2.6** An epistemic structure $(W, R, \models)$ is **induced** if

$$uRv \quad \Leftrightarrow \quad \textit{for all } F\ (u \models \mathbf{K}F \ \Rightarrow\ v \models F). \tag{3.5}$$

Informally, in an induced epistemic structure $(W, R, \models)$, the accessibility relation $R$ is the largest possible given $(W, \models)$. Since, in canonical models accessibility relation $R$ satisfies (3.5), all canonical models are induced epistemic structures.

## 3.3 On Induced Epistemic Structures

Let $(W, R, \models)$ be an induced epistemic structure over S5.

**Proposition 3.3.1** *$R$ is an equivalence relation on $W$.*

***Proof*** Reflexivity and transitivity are immediate. Checking symmetry is also standard. Let $wRx$, and suppose $x \models \mathbf{K}F$. We have to prove that $w \models F$. Suppose $w \not\models F$, then, by reflexivity in S5, $w \not\models \mathbf{K}F$, hence $w \models \neg\mathbf{K}F$. By negative introspection, $w \models \mathbf{K}\neg\mathbf{K}F$. By definition of $R$, $x \models \neg\mathbf{K}F$, which is impossible since $x$ is consistent and $x \models \mathbf{K}F$. □

The intuition of indistinguishability for states from $R(w)$ in epistemic structures is similar to Kripke models: we can interpret $R(w)$ as some set of states indistinguishable from $w$ by facts known to the agent. Apparently, $w \Vdash \mathbf{K}F$ yields $R(w) \Vdash F$ and all facts known at $w$ are true everywhere in $R(w)$. So, a state $x \in R(w)$ cannot be distinguished from $w$ by any fact known to the agent.

The principal difference between epistemic structures and Kripke models is that in the former, validity of $F$ in $R(w)$ does not yield knowledge of $F$: there is room for ignorance of agents about valid facts. In particular, it is possible to have $F$ throughout $R(w)$, but $\neg\mathbf{K}F$ at $w$.

The following proposition shows that knowledge assertions respect indistinguishability: either $\mathbf{K}F$ holds everywhere in $R(w)$, or $\neg\mathbf{K}F$ holds everywhere in $R(w)$.

**Corollary 3.3.2** *$R(w) \models \mathbf{K}F$ or $R(w) \models \neg\mathbf{K}F$.*

Indeed, suppose $w \models \mathbf{K}F$, but for some $x \in R(w)$, $x \models \neg\mathbf{K}F$. By negative introspection, $x \models \mathbf{K}\neg\mathbf{K}F$, hence $R(x) \models \neg\mathbf{K}F$. Since, by Proposition 3.3.1, $w \in R(x)$, $w \models \neg\mathbf{K}F$. A contradiction.

## 3.4 Canonical Structures for S5 with a Single Letter

In this section we will offer a useful elaborate example of canonical model constructions associated to S5 with a single propositional letter $p$, S5($p$). Such epistemic structures are defined by their possible worlds $W$ and truth relations $\models$ since the accessibility relations are induced and can be recovered from $(W, \models)$.

We first note that the modal-free fragment of S5($p$), i.e., the usual classical propositional logic with a single propositional letter $p$, admits two possible worlds: one generated by $\{p\}$ and the other generated by $\{\neg p\}$.

We claim that S5($p$) admits exactly four possible worlds (maximal consistent sets):

- $A$, generated by $\{\mathbf{K}p\}(= \{p, \mathbf{K}p\})$;
- $B$, generated by $\{p, \neg\mathbf{K}p\}$;
- $C$, generated by $\{\neg p, \neg\mathbf{K}\neg p\}$;
- $D$, generated by $\{\mathbf{K}\neg p\}(= \{\neg p, \mathbf{K}\neg p\})$.

**Consistency of each of** $A$–$D$ is straightforward since each has an easy Kripke model.

**Completeness of** $A$. First we note that $A$ is closed under Necessitation: $A \vdash F$ yields $A \vdash \mathbf{K}F$. Standard induction on derivations of $F$. The key point here is that $A \vdash \mathbf{K}A$. Once we establish Necessitation in $A$, we proceed to proving that for each $F$, $A \vdash F$ or $A \vdash \neg F$. Induction on $F$. Obvious for atomic formulas and Boolean connectives. Let $F = \mathbf{K}X$. If $A \vdash X$, then, by Necessitation, $A \vdash \mathbf{K}X$. If $A \vdash \neg X$, then, by reflexivity, $A \vdash \neg\mathbf{K}X$.

**Completeness of** $B$. Here Necessitation is not admissible since $B \vdash p$, but $B \nvdash \mathbf{K}p$. We will use the S5-normal forms, cf. Meyer and van der Hoek (2004).

**Lemma 3.4.1** (S5 normal forms) *In* S5*, every formula is provably equivalent to a formula in normal form which is a disjunction of conjunctions of type*

$$\alpha \wedge \mathbf{K}\beta \wedge \neg\mathbf{K}\gamma_1 \wedge \cdots \wedge \neg\mathbf{K}\gamma_m \tag{3.6}$$

*where* $\alpha, \beta, \gamma_1, \ldots, \gamma_m$ *are all purely propositional formulas. For* S5($p$) *we may assume that each of them is one of* $\top, \bot, p, \neg p$.

It now suffices to check that for each formula $F$ of type (3.6), $B \vdash F$ or $B \vdash \neg F$.

If $\alpha = \bot, \neg p$, then $B \vdash \neg F$. If $\alpha = \top, p$, then $B \vdash \alpha$ and we proceed to $\beta$.

If $\beta = \bot, \neg p$, then, by reflexivity, $B \vdash \neg F$. If $\beta = p$, then again, $B \vdash \neg F$. If $\beta = \top$, then $B \vdash \mathbf{K}\beta$ and we proceed to $\gamma_i$.

If at least one of $\gamma_i$ is $\top$, then $B \vdash \neg F$. Otherwise, all conjuncts in $F$ are provable in $B$. Indeed, for $\gamma_i = \bot$, use $B \vdash \neg\mathbf{K}\bot$. For $\gamma_i = p$ use the fact that $\neg\mathbf{K}p \in B$. For $\gamma_i = \neg p$, use reflexivity $p \to \neg\mathbf{K}\neg p$. In either case, $B \vdash \neg\mathbf{K}\gamma_i$.[5]

[5] A similar normal form-based proof of completeness can be given for each of $A$–$D$, but we have opted for Necessitation-based proof for $A$ and $D$ to underline the fact that both $A$ and $D$ enjoy Necessitation.

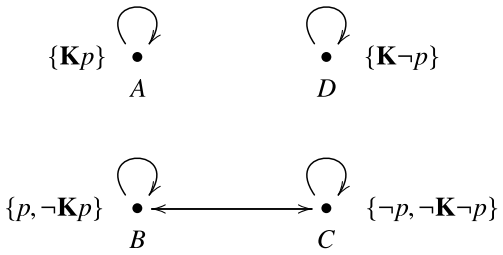



**Fig. 3.3** Canonical model for S5($p$)

**Completeness of** $C$. Similar to $B$

**Completeness of** $D$. Similar to $A$, since $D$ also enjoys Necessitation.

The collection of $A$–$D$ exhausts all combinatorial possibilities for states over S5($p$). Indeed, for the remaining four possibilities for $p$ and knowledge assertion about $p$, $\{p, \mathbf{K}\neg p\}$ and $\{\neg p, \mathbf{K}p\}$ are inconsistent and so are any of its extensions. The last two options: $\{p, \neg\mathbf{K}\neg p\} = \{p\}$ and $\{\neg p, \neg\mathbf{K}p\} = \{\neg p\}$, generate no new states. Indeed, for completion of $\{p\}$, we have to add either $\mathbf{K}p$ and get $A$ or $\neg\mathbf{K}p$ and get $B$. Likewise, any completion of $\{\neg p\}$ yields either $C$ or $D$.

Finally, we describe the induced accessibility relation $R$ on $W = \{A, B, C, D\}$. By Proposition 3.3.1, $R$ is an equivalence relation on $W$. Consider all six possible pairs of different states in $W$ and rule out the ones that are not accessible from each other.

By definition of $R$, for $ARX$, $p$ should be in $X$, which rules out $ARC$, $ARD$. Likewise, $B$ and $D$ are not related by $R$.

$\{A, B\}$ and $\{C, D\}$ are not connected due to positive introspection, e.g., since $\mathbf{K}p \in A$, $\mathbf{KK}p \in A$ too, hence $ARX$ yields $\mathbf{K}p \in X$; this rules out $ARB$.

The only remaining possibility for $R$-connection is pair $\{B, C\}$, and they are related! From the Kripke canonical model perspective, there should be a state in $W$ accessible from $B$ in which $\neg p$ holds, and $C$ is the only remaining possibility, hence $BRC$. The resulting picture of the canonical model for S5($p$) is (Fig. 3.3).

**Proposition 3.4.2** *Each of fifteen non-empty subsets of* $\{A, B, C, D\}$ *is an epistemic structure. The following seven of them are fully observable and hence Kripke models:*

$$\{A\},\ \{D\},\ \{A, D\},\ \{B, C\},\ \{A, B, C\},\ \{B, C, D\},\ \{A, B, C, D\}.$$

*The remaining eight are epistemic structures which are not fully observable and thus not Kripke models.*

Among epistemic structures here is $\{A, B\}$ which may be regarded as the canonical epistemic structure of $\Gamma = \{p\}$ and is therefore not a Kripke model.

## 3.5 From Epistemic Structures to Kripke Models

Each Kripke model is an epistemic structure; epistemic structures constitute a generalization of Kripke models that are able to account for more epistemic situations.

How do we build epistemic structures in general? To define an epistemic structure, we have to specify complete truth relations at each world. This is a task which Kripke models do naturally. Can we combine the generality of epistemic structures and the convenience of Kripke models?

In this section we show that each epistemic structure is a submodel of an appropriate Kripke model. This suggests a general method, *scaffolding*, of defining epistemic structures. This is the method: we first build a Kripke model, and then we carve from it the desired observable submodel.

**Theorem 3.5.1** *For any epistemic structure* $\mathcal{S} = (W, R, \models)$ *there is a Kripke model* $(\widetilde{W}, \widetilde{R}, \Vdash)$ *(which we call a "scaffolding of* $\mathcal{S}$*") such that*
*(a)* $W \subseteq \widetilde{W}$*;*
*(b)* $R \subseteq \widetilde{R}$*;*
*(c) for each* $u \in W$ *and each* $F$, $u \models F$ *iff* $u \Vdash F$.

***Proof*** Take

$$\mathcal{K} = (\widetilde{W}, \widetilde{R}, \Vdash)$$

to be the canonical Kripke model for the background modal logic.[6] Each $u \in W$ is also a world in $\mathcal{K}$ and this secures (a). A formula $F$ holds at $u$ iff $F \in u$: this yields (c). For (b), it suffices to note that $uRv$ yields $u\widetilde{R}v$. □

**Example 3.5.2** Here is a simple example of scaffolding. The aforementioned epistemic structure $\mathcal{M}_1$ from Example 3.1.1, Fig. 3.1, is naturally embedded into Kripke model $\mathcal{M}_2$. A singleton structure $\mathcal{M}_1$ coincides with world $u$ in $\mathcal{M}_2$ (Fig. 3.4).

We see here that the Kripke model requires two worlds to emulate a singleton epistemic structure.[7]

**Definition 3.5.3** Let $\Gamma$ be a set of formulas. By $\Gamma \models F$ we understand the situation when for each epistemic structure $\mathcal{S}$ and its state $u$,

$$\mathcal{S}, u \models \Gamma \quad \Rightarrow \quad \mathcal{S}, u \models F.$$

**Proposition 3.5.4** *Soundness and completeness w.r.t. epistemic structures:*

$$\Gamma \vdash F \text{ iff } \Gamma \models F.$$

[6] Which we assume to have the canonical model in the usual sense.

[7] It is easy to build a singleton epistemic structure for which any scaffolding Kripke model is infinite.

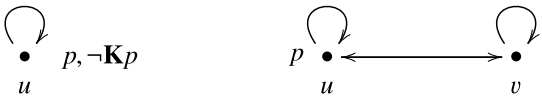



**Fig. 3.4** Epistemic structure $\mathcal{M}_1$ and Kripke model $\mathcal{M}_2$

***Proof*** Soundness. If $\Gamma \vdash F$, then $F$ belongs to any maximal consistent extension of $\Gamma$ and hence is true at each state of each epistemic structure of $\Gamma$.

Completeness is immediate from the saturation lemma: if $\Gamma \nvdash F$, then $\Gamma \cup \{\neg F\}$ is consistent and its maximal consistent extension $\Delta$ is a desired singleton epistemic structure: $\Delta \not\models F$. □

Theorem 3.5.1 suggests the following definition.

**Definition 3.5.5 An observable model** is a pair

$$O = (\mathcal{K}, S)$$

where $\mathcal{K}$ is a Kripke model and $S$ is an epistemic structure which is a submodel of $\mathcal{K}$ in the sense of Theorem 3.5.1. A formula $F$ holds at $u \in S$ in $(\mathcal{K}, S)$ if $u \Vdash F$ in $\mathcal{K}$.

An epistemic structure $S$ itself can be viewed as an observable model $(\mathcal{K}, S)$ with the canonical scaffolding model $\mathcal{K}$. A formula $F$ holds at $u \in S$ in $(\mathcal{K}, S)$, if $u \Vdash F$ in $\mathcal{K}$.

We see that observable models do not yield new entailment relation:

$$\Gamma \models F \textit{ in Kripke models} \quad \textit{iff} \quad \Gamma \models F \textit{ in observable models}.$$

Consequently, there has been no reason to invoke observable models for the traditional soundness and completeness studies of modal logic. However, epistemic contexts require better control over possible worlds and their ontological status. This is where observable models, which may be regarded as an epistemically sensitive version of Kripke models, may become useful.

Kripke models constitute a comprehensive semantic tool in modal logic: every consistent configuration is realized at an appropriate node of an appropriate model. However, in modal logic, we do not normally care how "possible" the states in a Kripke model are; their formal consistency is sufficient. Epistemic scenarios are different. We can imagine a situation in which some propositions $\Gamma$ (e.g., reflecting the state of nature) should hold at all possible states. Furthermore, if this $\Gamma$ is not common knowledge the corresponding set of states is, generally speaking, not a Kripke model, and requires a broader approach.

The standard Kripke modeling is a nondeterministic process which takes a syntactic description of an epistemic scenario $\mathcal{E}$ and figures out the structure of possible worlds and epistemic constraints given by $\mathcal{E}$. It then generates a "natural" Kripke model $\mathcal{K}$ which accommodates all these data thus tacitly assuming $\mathcal{K}$ to be common knowledge.

The observable model building (also nondeterministic) takes $\mathcal{E}$, finds a scaffolding Kripke model $\mathcal{K}$ for $\mathcal{E}$ (which technically can be the same "natural model") and, for a given agent, singles out a set of worlds and accessibilities $S$ in $\mathcal{K}$ which are explicit in $\mathcal{E}$. This $(\mathcal{K}, S)$ is our observable model. In Sect. 3.10.3 we briefly discuss constructive ways to build observable models based on partial descriptions.

## 3.6 Observable Models for Epistemic Modeling

Kripke models represent ignorance by building special "possible worlds": these worlds, however, can be epistemically impossible for the agent, cf. Examples 3.1.2, 3.1.3 and 3.5.2. In this respect, observable models offer a more refined foundational picture than Kripke models alone.

Another area in which observable models can have advantage is avoiding unnecessary *pro forma* worlds.

### *3.6.1 Ignorant Agent Example*

Here is a natural example, *Ignorant Agent*, of a two state observable structure for which scaffolding Kripke models are infinite and their building does not make much sense.

Consider S4 with a single propositional letter $p$ and a situation in which the agent has no specific knowledge about the state of nature, i.e., the agent knows only logical truths derived in S4. Consider two sets of formulas:

$$\Gamma^+ = \{p\} \cup \{\neg \mathbf{K} F \mid \mathsf{S4} \nvdash F\}, \quad \Gamma^- = \{\neg p\} \cup \{\neg \mathbf{K} F \mid \mathsf{S4} \nvdash F\}$$

**Proposition 3.6.1** *$\Gamma^+$ and $\Gamma^-$ are consistent, complete, and decidable.*

***Proof*** Consider $\Gamma^+$, the case of $\Gamma^-$ is similar.

**Consistency**. Suppose otherwise: $\Gamma^+ \vdash \bot$. Then

$$p, \neg \mathbf{K} F_1, \ldots, \neg \mathbf{K} F_n \vdash \bot$$

for some $F_i$'s not provable in S4. By S4-reasoning,

$$\mathsf{S4} \vdash p \to (\mathbf{K} F_1 \vee \cdots \vee \mathbf{K} F_n).$$

Since each of $F_i$'s has a countermodel, by a standard S4-model construction, we build an S4 Kripke model, at the root node $r$ of which $r \Vdash p$ and $r \nVdash \mathbf{K} F_1 \vee \cdots \vee \mathbf{K} F_n$, a contradiction.

**Fig. 3.5** Epistemic structure $\mathcal{M}_3$

**Completeness**: For each $F$, $\Gamma^+ \vdash F$ or $\Gamma^+ \vdash \neg F$. Induction on $F$. Atomic and Boolean cases are trivial. Let $F$ be $\mathbf{K}X$ for some $X$. Case 1. $\mathsf{S4} \vdash X$. Then $\mathsf{S4} \vdash \mathbf{K}X$, $\Gamma^+ \vdash \mathbf{K}X$. Case 2. $\mathsf{S4} \nvdash X$. Then $\Gamma^+ \vdash \neg\mathbf{K}X$.

**Decidability** of $\Gamma^+$ is immediate due to decidability of $\mathsf{S4}$. □

A natural epistemic structure for Ignorant Agent is $\mathcal{M}_3$ (Fig. 3.5):

It suffices to show that worlds $\Gamma^+$ and $\Gamma^-$ are indeed indistinguishable, i.e., $\Gamma^+ \vdash \mathbf{K}F$ yields $\Gamma^- \vdash F$ (and symmetrically for $\Gamma^-$ and $\Gamma^+$). Suppose $\Gamma^+ \vdash \mathbf{K}F$. Then $\mathsf{S4} \vdash F$, since otherwise $\mathsf{S4} \nvdash F$ and $\neg\mathbf{K}F \in \Gamma^+$ which would make $\Gamma^+$ inconsistent. Then $\Gamma^- \vdash F$.

Any scaffolding Kripke model for $\mathcal{M}_3$ is infinite. It must explicitly contain counter-worlds for all $F$'s s.t. $\mathsf{S4} \nvdash F$ which does not make much sense since $\mathsf{S4}$ is decidable and so there is no need to carry additional worlds in order to define a given $\mathsf{S4}$ situation epistemically.

## 3.7 Intuitionistic Observable Models

Nodes in an intuitionistic Kripke model

$$\mathcal{K} = (W, \preceq, \Vdash)$$

are information states in a discovery process and

$$u \preceq v$$

means that $v$ has more truths than $u$. In order to use $(W, \preceq, \Vdash)$ as a specification device, we have to assume the knowledge of the model by the agent. Without this assumption, the fact that $F$ does not hold at all states accessible from $w$ can be unknown and cannot constitute a sufficient ground for claiming $\neg F$ at $w$. Although

$$w \Vdash \neg F \quad \Rightarrow \quad v \nVdash F \textit{ for all } v s.t. w \preceq v,$$

the converse "$\Leftarrow$" does not seem well-principled: the fact that

$$v \nVdash F \textit{ for all } v \textit{ s.t. } w \preceq v$$

alone, without knowledge of the model, does not produce enough evidence at $w$ to conclude $\neg F$.

Consider a submodel $S = (W_S, \preceq_S, \models)$ of an intuitionistic Kripke model $\mathcal{K} = (W, \preceq, \Vdash)$ with

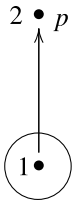
- $W_S \subseteq W$, states from $W_S$ are called "observable states";
- "observable accessibility" relation $\preceq_S$ on $W_S$ is a subset of $\preceq$;
- for $v \in W_S$, $v \models F$ is defined as $v \Vdash F$.

**An intuitionistic observable model** is, by definition, a pair

$$(\mathcal{K}, S).$$

Whereas the whole scenario can be represented traditionally, by a Kripke model $\mathcal{K}$, only a part $S$ of $\mathcal{K}$ is observable by the agent. Within the traditional Kripkean approach, we usually assume that a model $\mathcal{K}$ is known to the agent and recover the truths at each world. Within the observable intuitionistic model $(\mathcal{K}, S)$, we are given evaluations at each world of $S$ inherited from $\mathcal{K}$. A scaffolding Kripke model $\mathcal{K}$ is one of possible methods of defining evaluations in the worlds from $S$.

**Example 3.7.1** Intuitionistic Kripke model $\mathcal{K}$: $W = \{1, 2\}$, $1 \preceq 2$, $1 \nVdash p$, $2 \Vdash p$.




*Observable model* is $S$ (circled): $W_S = \{1\}$, $1 \preceq_S 1$, $1 \not\models p$. In $S$, $p$ does not hold, but yet $1 \nVdash \neg p$. $S$ itself is not a Kripke model!

Observable intuitionistic models can help to ease tension between Kripkean intuitionistic semantics and the intended Brouwer–Heyting–Kolmogorov, BHK, semantics of proofs for intuitionistic logic, cf. Artemov and Fitting (2019). The standard Kripkean condition

$$w \Vdash \neg F \quad \Leftrightarrow \quad v \nVdash F \text{ for all } v \text{ s.t. } w \preceq v \tag{3.7}$$

appears to require knowledge of the model, which is neither well-principled, nor realistic. Observable models do not assume knowledge of the whole model which rules out establishing $\neg F$ at $u$ in a non-BHK manner (3.7). An exploration of connections between observable models and BHK semantics appears to be an appealing research project.

## 3.8 Canonical Observable Models in a General Setting

In this section we tap the multi-agent case.

A generalization of observable models from a single agent setting to multiple agents requires many design decisions which should be properly motivated. We leave this for future studies. To give a flavor of what kind of new phenomena to expect, we consider here one meaningful case of canonical models in $\mathsf{S5}^n$.

**Definition 3.8.1** A *canonical observable model of a set of formulas* $\Gamma$, $CM(\Gamma)$, is, by definition, the collection of all maximal consistent sets containing $\Gamma$ with induced accessibility relations $R_1, R_2, \ldots, R_n$:

$$u R_i v \quad \textit{iff} \quad \textit{for all } F \ (u \models \mathbf{K}_i F \ \Rightarrow \ v \models F)$$

We show that for many (intuitively, most) $\Gamma$'s, the corresponding canonical model $CM(\Gamma)$ is not fully observable, hence not a Kripke model. We give a criterion for when the canonical model of $\Gamma$ is a Kripke model:

$$\Gamma \textit{ is closed under Necessitation}$$

or, equivalently,

$$\Gamma \textit{ proves its own common knowledge}.$$

**Example 3.8.2** Canonical model $CM(p)$ for $\Gamma = \{p\}$ in $\mathsf{S5}$ has been described in Sect. 3.4, Fig. 3.2. There are two possible worlds in $CM(p)$, generated by $\{\mathbf{K}p\}$ (world $A$) and $\{p, \neg\mathbf{K}p\}$ (world $B$). Worlds $A$ and $B$ are not connected by the indistinguishability relation $R$: $p$ holds at both worlds, but is not known in $B$. The canonical model $CM(p)$ is not a Kripke model, since $p$ holds in $CM(p)$, but $\mathbf{K}p$ does not.

### *3.8.1 Common Knowledge and Necessitation*

Consider a representative case of two agents. We will use the abbreviation for "everybody's knowledge"

$$\mathbf{E}X = \mathbf{K}_1 X \wedge \mathbf{K}_2 X,$$

and the abbreviation for "common knowledge"

$$\mathbf{C}X = \{X, \ \mathbf{E}X, \ \mathbf{E}^2 X, \ \mathbf{E}^3 X, \ \ldots\}.$$

As one can see, $\mathbf{C}X$ is an infinite (though easy to describe and decidable) set of formulas. Since modalities $\mathbf{K}_i$ commute with the conjunction $\wedge$, $\mathbf{C}X$ is logically

equivalent in $\mathsf{S5}^2$ to the set of all formulas which are $X$ prefixed by iterated knowledge modalities:

$$\{P_1 P_2 \dots P_k X \mid k = 0, 1, 2, \dots, \quad P_i \in \{\mathbf{K}_1, \mathbf{K}_2\}\}$$

and we regard this as an alternative definition of $\mathbf{C}X$. Naturally, the set of formulas

$$\mathbf{C}\Gamma = \bigcup\{\mathbf{C}F \mid F \in \Gamma\}$$

formally represents the property "$\Gamma$ is common knowledge."

The following proposition states that the rule of Necessitation corresponds to derivable common knowledge of assumptions.

**Proposition 3.8.3** *A set of formulas $\Gamma$ is closed under Necessitation if and only if $\Gamma \vdash \mathbf{C}\Gamma$, i.e., $\Gamma$ proves its own common knowledge.*

***Proof*** Direction 'if.' Assume $\Gamma \vdash \mathbf{C}\Gamma$ and prove by induction on derivations that $\Gamma \vdash X$ yields $\Gamma \vdash \mathbf{K}_i X$. For $X$ being a theorem of $\mathsf{S5}^2$, this follows from the rule of Necessitation in $\mathsf{S5}^2$. For $X \in \Gamma$, it follows from the assumption that $\Gamma \vdash \mathbf{C}X$, hence $\Gamma \vdash \mathbf{K}_i X$. If $X$ is obtained from *Modus Ponens*, $\Gamma \vdash Y \to X$ and $\Gamma \vdash Y$. By IH, $\Gamma \vdash \mathbf{K}_i(Y \to X)$ and $\Gamma \vdash \mathbf{K}_i Y$. By the distributivity principle of $\mathsf{S5}^2$, $\Gamma \vdash \mathbf{K}_i X$.

For 'only if,' suppose that $\Gamma$ is closed under Necessitation and $F \in \Gamma$, hence $\Gamma \vdash F$. Using appropriate instances of the Necessitation rule in $\Gamma$ we can derive $P_1 P_2 P_3, \dots, P_k F$ for each prefix $P_1 P_2 P_3, \dots, P_k$ with $P_i$ being one of $\mathbf{K}_1, \mathbf{K}_2$. Therefore, $\Gamma \vdash \mathbf{C}F$ for each $F \in \Gamma$, and $\Gamma \vdash \mathbf{C}\Gamma$. □

### 3.8.2 Canonical Models of Sets of Assumptions

We answer the question of when the canonical model of $\Gamma$ is a Kripke model.

**Theorem 3.8.4** *The following are equivalent:*
*(a) $CM(\Gamma)$ is fully observable (i.e., a Kripke model);*
*(b) $\Gamma$ admits Necessitation;*
*(c) $\Gamma$ proves its own common knowledge.*

***Proof*** The fact that (b) is equivalent to (c) has already been established in Proposition 3.8.3. We now check that (a) and (b) are equivalent.

If $\Gamma$ does not admit Necessitation, there is a formula $F$ such that $\Gamma \vdash F$, but $\Gamma \nvdash \mathbf{K}F$.[8] Hence $F$ holds everywhere in $CM(\Gamma)$ and $\neg\mathbf{K}F$ is consistent with $\Gamma$. Consider a maximal consistent extension $u$ of $\Gamma \cup \{\neg\mathbf{K}F\}$. Obviously, $\neg\mathbf{K}F$ holds in $u$ and $F$ holds in $R(u)$ which makes $CM(\Gamma)$ not fully observable and hence not a Kripke model.

[8] Here we ignore indices $i$ in $\mathbf{K}_i$ and $R_i$.

Suppose $\Gamma$ admits Necessitation. We have to prove that observable model $CM(\Gamma)$ with the canonical evaluation

$$u \models F \quad \Leftrightarrow \quad F \in u$$

and the induced $R$ is fully observable, i.e.

$$R(u) \models F \quad \Rightarrow \quad u \models \mathbf{K}F.$$

Following the standard construction, consider an arbitrary $u \in CM(\Gamma)$ and put

$$u^{\mathbf{K}} = \{F \mid \mathbf{K}F \in u\}.$$

Note that $\Gamma \subseteq u^{\mathbf{K}}$. Indeed, by Necessitation in $\Gamma$, $X \in \Gamma$ yields that $\mathbf{K}X \in \Gamma$ hence $\mathbf{K}X \in u$ and $X \in u^{\mathbf{K}}$. Suppose $u \not\models \mathbf{K}F$. Then $u^{\mathbf{K}} \cup \{\neg F\}$ is consistent. Indeed, otherwise

$$\Gamma \vdash X_1 \wedge \cdots \wedge X_n \to F$$

for some $X_i \in u^{\mathbf{K}}$. Since $\Gamma$ is closed under Necessitation,

$$\Gamma \vdash \mathbf{K}(X_1 \wedge \cdots \wedge X_n \to F).$$

By standard S5-reasoning,

$$\Gamma \vdash \mathbf{K}X_1 \wedge \cdots \wedge \mathbf{K}X_n \to \mathbf{K}F.$$

Since $\mathbf{K}X_i \in u$, $\mathbf{K}F$ should be in $u$ as well—a contradiction.

Now consider a maximal consistent set $v$ extending $u^{\mathbf{K}} \cup \{\neg F\}$. Since $\Gamma \subseteq u^{\mathbf{K}}$, $v \in CM(\Gamma)$. Since $u^{\mathbf{K}} \subseteq v$, $uRv$. Since $\neg F \in v$, $F \notin v$ hence $v \not\models F$. □

**Example 3.8.5** Consider the canonical model for S5$(p)$, Sect. 3.4. The canonical model of $\Gamma = \{p\}$, $CM(p)$,[9] is the set of worlds at which $p$ holds, i.e., $W = \{A, B\}$. Model $CM(p)$ is not fully observable, not a Kripke model, and is an illustration of Theorem 3.8.4, since $\{p\}$ is not closed under Necessitation (Fig. 3.6).

The canonical model of $\Gamma = \{\mathbf{K}p\}$ is the set of worlds at which $\mathbf{K}p$ holds, i.e., $W = \{A\}$. This $\Gamma$ enjoys Necessitation, its canonical observable model is fully observable, i.e., is a Kripke model.

The canonical model of $\Gamma = \{\neg\mathbf{K}p, \neg\mathbf{K}\neg p\}$. By negative and positive introspection, $\Gamma$ is closed under Necessitation, hence $CM(\Gamma)$ should also be fully observable. Worlds $A$ and $D$ are not compatible with $\Gamma$ and hence are not in $CM(\Gamma)$. Since neither $B$ nor $C$ is fully observable, $CM(\Gamma) = \{B, C\}$. It is an easy exercise to derive $\neg\mathbf{K}p$ in $C$ and $\neg\mathbf{K}\neg p$ in $B$. This is a Kripkean situation, i.e., $CM(\{\neg\mathbf{K}p, \neg\mathbf{K}\neg p\})$ is fully observable.

[9] We drop brackets in $CM(\{p\})$ and similar cases for better readability.

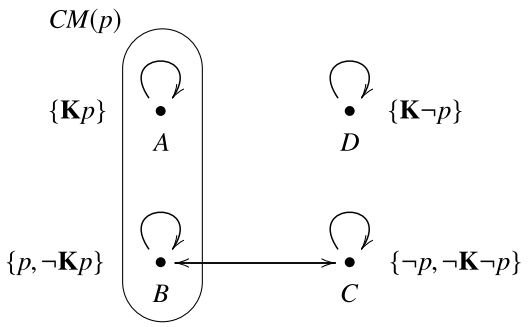



**Fig. 3.6** Canonical model $CM(p)$ (in the oval)

## 3.9 Findings and Suggestions

The principal contributions of this paper are conceptual: we suggest streamlining the foundations of epistemic modeling.

When epistemic logic starts from a given set of possible worlds and defines an accessibility relation in the standard way, the result might not be a Kripke model: only those structures that are fully observable are Kripke models. The fully observable property is a propositional equivalent to common knowledge of the model, which is quite restrictive but has been tacitly assumed in epistemic modeling. This assumption should be acknowledged, made explicit, and properly studied.

We have sketched a basic theory of observable models without the knowledge of the model constraint. This is a step towards a general theory of observable modeling with possible worlds; more studies are needed to capture partial and asymmetric knowledge of the model for multiple agents.

A similar suggestion is made for intuitionistic models. A well-known disconnection between intuitionistic Kripke models and the intended constructive semantics of intuitionistic logic can be mitigated by viewing intuitionistic possible worlds models as intuitionistic observable models.

## 3.10 Discussion

Observable models are not motivated by liberalizing the condition for ignorance *per se*, but rather are the result of a direct formalization of the standard epistemic desiderata for possible world semantics (minus the unwarranted *CKM* requirement). Naturally, observable models constitute a broader than Kripke's class of epistemic models.

Theorem 3.5.1 connects observable models to the Kripke framework: each epistemic structure is a submodel of an appropriate Kripke model $\mathcal{K}$ with the evaluations inherited from $\mathcal{K}$. Whereas in the traditional approach, to build a possible worlds model of an epistemic situation $\mathcal{E}$ means to find an appropriate Kripke model, in

**Fig. 3.7** Kripke model $\mathcal{M}_2$

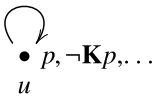



**Fig. 3.8** Epistemic structure $S$

our approach, we have to specify a scaffolding Kripke model and its observable submodel.

### *3.10.1 Solving the Puzzle from Example 3.1.2*

As we can see from Example 3.1.2, an S4-agent cannot know the model $\mathcal{M}_2$, Fig. 3.7. However, $\mathcal{M}_2$ can be regarded as a scaffolding Kripke model $\mathcal{K}$ in the observable model $(\mathcal{K}, S)$ with $S$ being its submodel, Fig. 3.8.

Evaluations of formulas at $u$ in $S$ are inherited from $\mathcal{K}$. Note that $S$ is not itself a Kripke model. Since the agent does not know the scaffolding model $\mathcal{K}$, the paradox disappears.

### *3.10.2 On Epistemic Modeling Without CKM*

A meaningful analysis of epistemic models with some limitations on the *CKM* requirement was made in Aucher (2010) that considers epistemic situations with different models for each agent. From the foundational perspective, this analysis goes halfway in the right direction since it does not question the *CKM* requirement for the base case of individual models. As the result, individual epistemic models in Aucher (2010) are traditional fully observable Kripke models. In the current paper we analyze epistemic models without *CKM* for individual knowers: a non-*CKM* individual epistemic model consists of two components

$$(\mathcal{K}, S)$$

with $\mathcal{K}$ being a traditional Kripke model, not necessarily having epistemic interpretation, and its epistemically meaningful submodel $S$ which is not necessarily a Kripke model.

What is the added value of $S$ if the scaffolding Kripke model $\mathcal{K}$ completely determines the values of all formulas in the real world?

Our answer: *the epistemic status of possible worlds and accessibility relations is itself a valuable epistemic information*. This is because an epistemic scenario is a simultaneous picture of real world and related epistemically possible worlds. In this particular sense, a scaffolding Kripke model alone may provide misleading expectations and answers by assuming that **all** scaffolding worlds are epistemically meaningful and known to the agent.

### *3.10.3 Observable Models and Partial Descriptions*

Modeling a syntactic description $\mathcal{D}$ by a single model $\mathcal{M}$ leads to overspecification; a typical example is Peano Arithmetic PA (an incomplete theory) and its standard model $\omega$ which is a complete overspecification of PA since each arithmetical sentence is either true or false in $\omega$.

There are several avenues of mitigating such overspecification including partial models with truth value gaps and superevaluations, in which a syntactic description $\mathcal{D}$ generates a class of models and a sentence $F$ holds in $\mathcal{D}$ iff $F$ is true in each of these models.[10]

Developing a comprehensive theory of observable models and their partial descriptions (epistemic theories) along the aforementioned ideas is a natural next step in this area.

## References

Artemov, S. (2016). *Knowing the model.* arXiv:1610.04955

Artemov, S. (2020). Observable models. In S. Artemov & A. Nerode (Eds.), *Logical foundations of computer science, LFCS 2020* (pp. 12–26). Cham: Springer.

Artemov, S., & Fitting, M. (2019). *Justification logic: Reasoning with reasons*. Cambridge: Cambridge University Press.

Aucher, G. (2010). An internal version of epistemic logic. *Studia Logica, 94*(1), 1–22.

Aumann, R. (1995). Backward induction and common knowledge of rationality. *Games and Economic Behavior, 8*(1), 6–19.

Fagin, R., Halpern, J. Y., Moses, Y., & Vardi, M. Y. (1995). *Reasoning about knowledge*. Cambridge: MIT Press.

Liu, Y. (2019). Two tales of epistemic models. *Thought: A Journal of Philosophy, 8*(4), 291–302.

Meyer, J.-J.C., & van der Hoek, W. (2004). *Epistemic logic for AI and computer science*. Cambridge: Cambridge University Press.

Williamson, T. (2015). A note on Gettier cases in epistemic logic. *Philosophical Studies, 172*(1), 129–140.

[10] In the standard curriculum of mathematical logic, the superevaluation approach is more popular, cf. Gödel's Completeness Theorem for the first-order logic.

# Chapter 4
# The Logic of Logical Necessity

**Andrew Bacon and Kit Fine**

**Abstract** Prior to Kripke's seminal work on the semantics of modal logic, McKinsey offered an alternative interpretation of the necessity operator, inspired by the Bolzano–Tarski notion of logical truth. According to this interpretation, 'it is necessary that A' is true just in case every sentence with the same logical form as A is true. In our paper, we investigate this interpretation of the modal operator, resolving some technical questions, and relating it to the logical interpretation of modality and some views in modal metaphysics. In particular, we present an hitherto unpublished solution to problems 41 and 42 from Friedman's 102 problems, which uses a different method of proof from the solution presented in the paper of Tadeusz Prucnal.

**Keywords** Logical necessity · Logical truth · McKinsey · Modality · Substitution · Tarski

## 4.1 Introduction

A common conception of a logical truth, often credited to Bolzano, is that of a sentence true in virtue of its logical form alone. In a given interpreted language one might make this precise by stipulating a sentence to be logically true if and only if the result of uniformly substituting any of the non-logical constants with expressions of the same grammatical category is true, and dually, logically consistent if and only if some substitution instance is true. For instance, 'if John is tall then John is tall' is a logical truth, since the result of substituting any name and predicate for

A. Bacon
USC School of Philosophy, 3709 Trousdale Pkwy., MHP 113, Los Angeles, CA 90089, USA
e-mail: a.bacon@usc.edu

K. Fine (✉)
Department of Philosophy, 5 Washington Place, New York, NY 10003, USA
e-mail: kf14@nyu.edu

© The Author(s), under exclusive license to Springer Nature Switzerland AG 2024

Y. Weiss and R. Birman (eds.), *Saul Kripke on Modal Logic*, Outstanding Contributions to Logic 30, https://doi.org/10.1007/978-3-031-57635-5_4

'John' and 'is tall' respectively results in a truth, whereas 'John is tall' is not a logical truth because it is either already false or the result of substituting the predicate 'not tall' for 'tall' in it is false.

This analysis makes salient a formal analogy between the notion of a substitution in the Bolzanoean definition of logical truth and logical consistency, and the notion of a possible world in the analysis of necessity and possibility in a Kripke model. Indeed, prior to Kripke's work on the semantics of modal logic, McKinsey proposed a substitutional interpretation—or more accurately, a constraint on such an interpretation—of the modal operator exactly along Bolzanoean lines.[1] Given a language $\mathcal{L}$ containing the usual truth functional connectives and a unary connective $\Box$, McKinsey laid out some conditions that the set of true sentences of this language should satisfy, on the intended interpretation of $\Box$. Apart from the usual clauses for the truth functional compounds, McKinsey requires that for every sentence $A$:

The Substitutional Constraint $\Box A$ is true if and only if $iA$ is true for every substitution $i$ of the language.

McKinsey leaves it open what sort of language $\mathcal{L}$ is, and the exact nature of the class of substitutions, stipulating only that the sentences of the language be closed under the truth-functional and modal connectives, and that the substitutions be closed under composition and contain a trivial substitution. We will consider more concrete versions of the substitutional constraint shortly.

Under an alternative approach, the logical truths are not as Bolzano thought of them, but are nonetheless given by some theory $\Delta$. Since $\Delta$ is being informally understood as the set of logical *truths* we should require that $\Delta$ be a subset of the truths, and one could then replace McKinsey's constraint with:

The Metalogical Constraint $\Box A$ is true if and only if $A \in \Delta$.

Constraining the interpretation of the modal operator $\Box$ to include in its extension propositions expressed by sentences in $\Delta$ and to exclude propositions expressed by sentences not in $\Delta$. Provided the theory $\Delta$ is closed under the rule of substitution, one direction of McKinsey's constraint is ensured, although not necessarily the other. In the case where $\mathcal{L}$ is the language of propositional modal logic, and $\Delta$ a logic (which we may assume to be at least closed under modus ponens and the rule of necessitation), we obtain an interpretation of propositional modal logic hit upon independently by Meyer and Fine in the 70s.[2]

---

[1] See McKinsey (1945).

[2] Specifically, the metalogical constraint is satisfied by any metavaluation (see Meyer, 1971, and Sect. 4.6.2 below): a valuation mapping sentences of propositional modal logic to 1 or 0, satisfying the conditions that $v(A \wedge B) = min(v(A), v(B))$, $v(\neg A) = 1 - v(A)$, and $v(\Box A) = 1$ iff $A \in \Delta$. If, additionally, $v(A) = 1$ whenever $A \in \Delta$ for any metavaluation $v$, $\Delta$ is called *coherent*. Fine (n.d.) applies the metalogical interpretation of $\Box$ to obtain simple proofs of the disjunction problem in modal and intuitionistic logics. For instance, any coherent modal logic, $\Delta$, will have the disjunction property, for if $\Box A_1 \vee \cdots \vee \Box A_n \in \Delta$, then $v(\Box A_1 \vee \cdots \vee \Box A_n) = 1$ is true in any metavaluation $v$ based on $\Delta$, and so $A_i \in \Delta$ for some $i$. See Sect. 4.6.2.

Whether or not these constraints are plausible will depend both on the interpretation of the modal operator and on the range of interpretations of the non-logical constants of the language in which the constraints are formulated.

Under what interpretations of the modal operator could these schemas be plausible? One possibility, following Quine, is to treat the modal operator as 'crypto-quotational', so that a sentence embedded under a modal operator should be understood as residing inside invisible quotation marks, and the modality ascribes some metalinguistic property to the embedded sentence. But although we motivated the constraints by analogy with the metalinguistic notion of logical truth, one doesn't *have* to identify the modality under either constraint with a metalinguistic property. One could instead posit a genuine propositional operator that yields an interpretation of $\Box$ under which the substitutional or metalogical conditions are satisfied with respect to a suitable language $\mathcal{L}$. The schemas then impose a substantive constraint on a propositional operator by articulating a sense in which the postulated notion of logical necessity stands to the world as logical truth stands to language, without thereby identifying the two.

The constraints are not plausible on arbitrary interpretations of the non-logical constants either. For instance, if the language in question contained predicates 'bachelor' and 'married' with their customary meanings then either constraint will imply, given reasonable side conditions, that it is possible that there are married bachelors.[3] While this result might be acceptable when it is interpreted as a metalinguistic logical consistency claim in disguise, it is arguably not so on the worldly interpretation proposed above. For plausibly, to be a bachelor *just is* to be a man who is not married, so, by Leibniz's law, the alleged possibility would imply that it is possible that there are married men who are not married, which is certainly not true on any candidate interpretation of the modal operator. (Such an appeal to Leibniz's law would not be legitimate on a crypto-quotational reading of the modal operator as logical consistency, since Leibniz's law does not permit the substitution of identicals within quotation marks.)

One can avoid these untoward results by working in what Russell calls a *logically perfect* language: a language where the non-logical constants do not denote logically complex properties and no two non-logical constants codenote.[4] The former requirement rules out predicates like 'bachelor' that denote conjunctions of simpler properties, and the latter pairs of synonymous predicates like 'lawyer' and 'attorney'.

---

[3] We can assume $\Delta$ is a *logical* theory, so it will not prove specific relationships concerning the non-logical constants, and we can also assume that the substitutions are rich enough we may substitute for 'married' and 'bachelor' predicates, like 'married' and 'not a bachelor', that are coinstantiated.

[4] Russell (1940, p. 25) writes: 'In a logically perfect language the words in a proposition would correspond one by one with the components of the corresponding fact, with the exception of such words as "or", "not", "if", "then", which have a different function. In a logically perfect language, there will be one word and no more for every simple object, and everything that is not simple will be expressed by a combination of words, by a combination derived, of course, from the words for the simple things that enter in, one word for each simple component.'

The schemas then capture a sort of Humean vision in which the logically simple, or fundamental, properties and relations can, in the proposed sense of 'can', stand in any consistent logical relationship to one another.

Lastly, McKinsey left it open what sort of language could be plugged into his constraint. It could simply be the language of propositional modal logic—and this will be the option we will explore in this paper—although it could be a more expressive language, such as a first-order or higher-order language. If it is a higher-order language there is the possibility of a Tarskian analysis of logical truth. A sentence $A(c_1 \ldots c_n)$ in non-logical constants $c_1 \ldots c_n$ is a logical truth, according to Tarski, when it is true under every possible interpretation of $c_1 \ldots c_n$. We can achieve generality over the interpretations $c_1 \ldots c_n$ in a higher-order language by simply quantifying into the positions they occupy, leaving the logical truth of $A(c_1 \ldots c_n)$ amounting to the truth of $\forall x_1 \ldots x_n . A(x_1 \ldots x_n)$, where each $x_i$ is a variable of the same type as $c_i$. Thus, for instance, the logical truth of 'If John is tall then John is tall' amounts to the mere truth of 'for any individual $x$ and property $Y$ if $x$ is $Y$ then $x$ is $Y$'. In this case, we can fully internalize what in the first two constraints was stated in metalinguistic terms, since $\forall x_1 \ldots x_n . A(x_1 \ldots x_n)$ is a sentence of the object language. So another constraint, in the same spirit as the substitutional and metalogical constraints, replaces the Bolzanoean conception of logical truth with the Tarskian, yielding:

The Tarskian Constraint $\quad \Box A(c_1 \ldots c_n) \leftrightarrow \forall x_1 \ldots x_n . A(x_1 \ldots x_n)$.

And this can likewise be thought of as articulating a Humean metaphysics of freely recombinable fundamental entities.[5]

The paper is organized as follows. In Sect. 4.2, McKinsey's substitutional constraint on the interpretation of necessity is made precise within the context of propositional modal logic, and associated notions of valuation and validity are defined. In Sect. 4.3 we raise the question of when an assignment of truth values to the sentence letters can be extended uniquely to a valuation of the modal language satisfying the substitutional constraint. We show that unique extensions exist for any class of non-modal substitutions, and for the full substitution class we present a previously unpublished proof of the existence of an extension, leaving the uniqueness as a conjecture. Sections 4.4 and 4.5 concern the logic of logical necessity. We consider the validity of various principles of modal logic with respect to various substitution classes, such as the McKinsey and Grzegorczyk axioms, and settle some questions raised in Humberstone (2016). Section 4.6 treats the metalogical and Tarskian constraints, and the paper concludes with some remarks on the substitutional approach to modal predicate logic in Sect. 4.7.

[5] See Bacon (2020).

## 4.2 Substitutional Interpretations of $\Box$

### *4.2.1 Preliminaries*

McKinsey does not specify the language he is working in or provide a concrete account of what he means by a substitution. Rather, he proceeds by imposing some abstract constraints on the language and the substitutions. He assumes that the language is closed under the formation of truth-functional compounds and necessity sentences. For simplicity, we will assume that our own language $\mathcal{L}$ is at least closed under conjunction and negation and the formation of necessity sentences:

L1. If $A, B \in \mathcal{L}$ then $(A \wedge B) \in \mathcal{L}$
L2. If $A \in \mathcal{L}$ then $\neg A \in \mathcal{L}$
L3. If $A \in \mathcal{L}$ then $\Box A \in \mathcal{L}$.

We will regiment McKinsey's account of substitution in terms of a set, $S$, of abstract substitutions and an action $\mu : S \times \mathcal{L} \to \mathcal{L}$ where $\mu(i, A)$—written as $iA$—informally represents the result of applying the substitution $i \in S$ to the sentence $A$ of $\mathcal{L}$ to produce another sentence of $\mathcal{L}$. We shall say that a pair $(S, \mu)$ is a *substitution class* if and only if the conditions below of Commutativity, Identity and Composition are satisfied. We will often suppress mention of the action $\mu$, and refer to a substitution class solely by its associated set of substitutions provided no ambiguity arises.

Commutativity The action of each substitution $i \in S$ on $\mathcal{L}$ commutes with the logical connectives, $\wedge, \neg, \Box$:

1. $i(A \wedge B) = (iA \wedge iB)$
2. $i\neg A = \neg iA$
3. $i\Box A = \Box iA$.

Identity There is a substitution $i \in S$ such that $iA = A$ for every sentence $A$ of $\mathcal{L}$.

Composition For any two substitutions, $i, j \in S$ there is as substitution $k \in S$ such that $kA = i(jA)$ for every sentence $A$ of $\mathcal{L}$.

As previously mentioned, these constraints may be satisfied by some very expressive languages, including first and higher-order languages. However we will mostly restrict our attention to propositional languages. Going forward, we will use $\mathcal{L}$ to refer to the propositional modal language with letters $p_0, p_1, p_2, \ldots$ and connectives $\wedge, \neg, \Box$. For technical reasons, it will sometimes be convenient to suppose that the language contains primitive 0-ary constants $\top$ and $\bot$, in which case the Commutativity condition must be extended to include the equations $i\top = \top$ and $i\bot = \bot$ for any $i \in S$. (Note that these equations might not hold if $\bot$, for instance, were treated as a defined connective, such as $(p_0 \wedge \neg p_0)$.)

Occasionally we will need to talk about other languages with further or fewer connectives. For any subset $\{C_1 \ldots C_n\} \subseteq \{\wedge, \vee, \neg, \to, \top, \bot, \Diamond, \Box\}$ we will write

$\mathcal{L}(C_1 \ldots C_n)$ to represent the sublanguage in the connectives $C_1 \ldots C_n$: the smallest set containing the sentence letters, and containing $C_m A_1 \ldots A_k$ whenever it contains $A_1 \ldots A_k, k$ the arity of $C_m$ and $1 \leq m \leq n$. $\mathcal{L}()$ thus refers to the set of sentence letters. A substitution class for a language with some of these logical connectives is defined as above except that we require the substitutions commute with any additional logical connectives, as with $i(A \vee B) = (iA \vee iB)$ or $i\Diamond A = \Diamond iA$. We say that a language is *complete* iff it is truth-functionally complete and contains either $\Box$ or $\Diamond$.

**Example 4.2.1** (*The full monoid of substitutions of propositional modal logic*) Let $S$ be the set of concrete substitutions of $\mathcal{L}$, i.e., functions $i : \mathcal{L}() \rightarrow \mathcal{L}$ from sentence letters to arbitrary sentences of $\mathcal{L}$. $\mu(i, A)$ may be defined as the result of uniformly substituting $i(p_k)$ for $p_k$ in $A$, for each $k \in \mathbb{N}$.

Commutativity and L1–L3 are clearly satisfied. Identity is witnessed by the element that maps each letter to itself, and composition by the element of $S$ that maps each letter $p_k$ to $\mu(i, j(p_k))$.

Note, however, that a set satisfying McKinsey's requirements may not be isomorphic to the full monoid of substitutions of a language. There might be distinct substitutions $i, j \in S$ such that $iA = jA$ for every sentence $A \in \mathcal{L}$ (there could, for instance, be two identity elements satisfying Identity). Say that $i \sim j$ iff $iA = jA$ for every $A \in \mathcal{L}$. If one quotients a set $S$ satisfying Commutativity, Identity and Composition by this equivalence relation one gets another set $S/\sim$ which satisfies Commutativity, Identity and Composition under the action defined by setting $\mu_\sim([i]_\sim, A) = \mu(i, A)$. Indeed, under this quotienting operation $S$ forms a monoid: the unit $\iota$ may be defined as the equivalence class $[i]_\sim$ of any element of $S$ satisfying Identity, and $[i] \circ [j]$ may defined to be the equivalence class of any $k$ satisfying Composition. For most purposes we can simply treat $S$ as a set of concrete substitutions of the language containing the identity substitution and closed under composition of substitutions. However, even if we impose these additional conditions, one still does not get that any *function* from the non-logical constants (the sentence letters in this case) to arbitrary expressions (sentences in this case) extends to a substitution in $S$. This property is distinctive to the full substitution class alone.

**Example 4.2.2** (*Substitutions within a sublanguage*) Write $S(C_1 \ldots C_n)$ for the substitution class defined by the set of functions

- $i : \mathcal{L}() \rightarrow \mathcal{L}(C_1 \ldots C_n)$

When $\mathcal{L}'$ is a language containing $C_1 \ldots C_n$ (so that $\mathcal{L}(C_1 \ldots C_n) \subseteq \mathcal{L}'$), the action of $i \in S(C_1 \ldots C_n)$ may be defined in the usual way (as in Example 4.2.1), thereby satisfying Commutativity.

Since $\mathcal{L}(C_1 \ldots C_n)$ contains each sentence letter, $S(C_1 \ldots C_n)$ contains the identity substitution and satisfies Identity. Since $\mathcal{L}(C_1 \ldots C_n)$ is itself a language, it is closed under substitutions of letters by sentences in $\mathcal{L}(C_1 \ldots C_n)$. So Composition is also satisfied.

We will also investigate a class of substitutions introduced in Humberstone (2016), in relation to McKinsey's theory of necessity:

**Example 4.2.3** (*Humberstone substitutions*) A Humberstone substitution is a function $i : \mathcal{L}() \to \mathcal{L}(\top\bot)$ such that

- $i(p_k)$ is either $p_k$, $\top$ or $\bot$.

These substitutions act on the language with primitive $\top$ and $\bot$ connectives (i.e., on $\mathcal{L}(\neg \wedge \Box\top\bot)$). $\mu(i, A)$ is defined in the usual way, so that McKinsey's conditions are satisfied. We call the set of Humberstone substitutions $H$.

A positive formula is one whose letters all occur positively, where the letters occurring positively and negatively in a formula are defined by a simultaneous recursion as follows:

- $P(p_k) = \{p_k\}$, $N(p_k) = \emptyset$,
- $P(A \wedge B) = P(A) \cup P(B)$, $N(A \wedge B) = N(A) \cup N(B)$,
- $P(\neg A) = N(A)$, $N(\neg A) = P(A)$
- $P(\Box A) = P(A)$, $N(\Box A) = N(A)$.

**Example 4.2.4** (*Positive substitutions*) A positive substitution is a substitution that maps letters to positive formulas, and acts on formulas in the usual way. We will call the class of positive substitutions $P$.

Lastly, we consider a special sort of substitution that will play a role later in our discussion of Carnap's theory of logical necessity.

**Example 4.2.5** (*Carnapian substitutions*) A Carnapian substitution is a function $i$ on sentence letters such that

- $i(p_k)$ is the result of prefixing some number (possibly zero) of $\neg$ signs to $p_k$.

and $\mu(i, A)$ is defined as before. We call the set of Carnapian substitutions $K$.

### *4.2.2 S-Valuations, Pre-validity and Validity*

We introduce the notion of an $S$-valuation: a valuation, taking sentences of the language $\mathcal{L}$ to truth values, that satisfies McKinsey's constraints.

**Definition 4.2.6** Suppose that $\mathcal{L}$, $S$ and $\mu$ satisfy L1–L3, Commutativity, Identity and Composition. An $S$-valuation is a function $v : \mathcal{L} \to \{0, 1\}$ such that

- $v(A \wedge B) = min(v(A), v(B))$
- $v(\neg A) = 1 - v(A)$
- $v(\Box A) = 1$ if and only if, for every $i \in S$, $v(iA) = 1$

By a *truth-value assignment* we mean a function defined on the sentence letters, $v^- : \mathcal{L}() \to \{0, 1\}$. A valuation $v$ extends a truth-value assignment, $v^-$, iff $v \restriction_{\mathcal{L}()} = v^-$.

Formally speaking we identify truth and falsity with the numbers 1 and 0, but we shall sometimes talk of a sentence being true or false in a valuation rather than receiving the value 1 or 0.

In cases where some of $\top$, $\bot$, $\vee$ or $\Diamond$ are also taken as primitive, the notion of valuation should respect the natural clauses for those connectives:

- $v(A \vee B) = max(v(A), v(B))$
- $v(\top) = 1$
- $v(\bot) = 0$
- $v(\Diamond A) = 1$ if and only if, for some $i \in S$, $v(i\,A) = 1$.

We introduce two senses in which a formula can be valid relative to a substitution class. The first is not so straightforwardly related to the Bolzano–Tarski conception of validity, since the set of valid sentences in this sense need not be closed under the rule of substitution.

**Definition 4.2.7** (*Pre-validity*) A sentence $A$ is pre-valid relative to the substitution class $S$ if and only if $v(A) = 1$ for every $S$-valuation $v$.

To see that pre-validity is not closed under the rule of substitution, consider any substitution class $S$ which contains a substitution $i$ such that $i(p_0) = \top$ (as with $S(C_1 \ldots C_n)$ for any set of connectives $C_1 \ldots C_n$ containing $\top$). For any $S$-valuation $v$, $v(\Diamond p_0) = 1$ because $v(ip_0) = v(\top) = 1$. So $\Diamond p_0$ is pre-valid relative to the class $S$. However the substitution instance $\Diamond\bot$ is not pre-valid when $\bot$ is in the language; indeed it is false in every $S$-valuation.

Validity proper thus cannot be understood as the result of closing the pre-validities under the rule of substitution. Rather we consider a sentence to be valid relative to a substitution class $S$ only if all of its substitution instances are pre-valid.

**Definition 4.2.8** (*Validity)* A sentence $A$ is valid relative to the substitution class $S$ if and only if $i\,A$ is pre-valid relative to the substitution class $S$, for every substitution $i$ in the full substitution class $S(\neg \wedge \Box)$.

If $S$ is a substitution class, then we will write $L(S)$ for the set of sentences valid with respect to $S$, and we call it the *logic of* $S$.

Note that the 'external' notion of validity defined above, for which we required the theorems of the logic to be valid, can be distinguished from the 'internal' notion of validity by which $\Box$ is interpreted. The former requires truth under all substitutions, whereas the latter requires only truth under all substitutions in $S$. Consequently, they can diverge when $\Box$ is interpreted by a notion of validity more restricted than our external notion of logical validity. According to the interpretation of $\Box$ as *logical* validity, however, the two notions will coincide, and for this reason the case where $S$ is the full substitution class is of special interest.[6]

[6] It is also possible to adopt an external notion of validity intermediate between prevalidity and validity, whose validities are those sentences with prevalid $S$-substitution instances. This secures the coincidence between the external and internal notions in another way, although this is not an avenue we have pursued. The rationale behind requiring logic to be closed under arbitrary substitutions, at

Let us remark briefly on the choice of primitives. It is easily verified that standard definitions of arbitrary truth-functional compounds in terms of $\neg$ and $\wedge$ receive the same truth value in any valuation. Similarly $\Diamond$ may be defined, as usual, in terms of $\neg$ and $\Box$. In most contexts it is more economical to work in the language $\mathcal{L}$, whose primitives consist of only $\wedge$, $\neg$ and $\Box$. But in some cases, the presence of the other primitives make the definitions simpler: for instance if we identified $\top$ and $\bot$ with particular definitions, such as $\neg(p_0 \wedge \neg p_0)$ and $p_0 \wedge \neg p_0$, then our official definition of $S(\top\bot)$ would not satisfy the commutativity equations $i\bot = \bot$ and $i\top = \top$.

The differences between the logics of different substitution classes will be the subject of the next few sections. However, we are already in a position to see how different substitution classes may behave differently at the level of pre-validity. We have, for instance, the following contrast between full and non-modal substitutions:

**Proposition 4.2.9** *The sentence $\Diamond(\Diamond p \wedge \Diamond\neg p \wedge \Box(p \rightarrow \Box p))$ (for $p$ a sentence letter) is pre-valid over $S(\neg \wedge \Box)$, while its negation is pre-valid over $S(\neg\wedge)$.*

***Proof*** Let $v$ be any $S(\neg \wedge \Box)$ valuation. First note that $v(\Box\top) = 1$ since $v(i\top) = v(\top) = 1$ for every $i \in S(\neg \wedge \Box)$. Similarly for any sentence letter $q$, $v(\neg\Box q) = 1$: $v(\Box q) = 0$ since $v(r \wedge \neg r) = 0$ and for some $i \in S(\neg \wedge \Box)$, $iq = r \wedge \neg r$.

We shall show that upon substituting $\Box q$ for $p$ in $(\Diamond p \wedge \Diamond\neg p \wedge \Box(p \rightarrow \Box p))$ we get something true in $v$, and so $\Box q$ witnesses the truth of $\Diamond(\Diamond p \wedge \Diamond\neg p \wedge \Box(p \rightarrow \Box p))$. (i) $v(\Diamond\Box q) = 1$ since $\Box q$ has a true-in-$v$ substitution, namely $v(\Box\top) = 1$. (ii) $v(\Diamond\neg\Box q) = 1$ since $\neg\Box q$ is already true in $v$. (iii) $v(\Box(\Box q \rightarrow \Box\Box q)) = 1$. Take any $i \in S(\neg \wedge \Box)$. If $v(\Box iq) = 1$ then $v(jiq) = 1$ for every $j$. This means $v(kliq) = 1$ for any $k$ and $l$, and so $v(\Box\Box iq) = 1$. Thus $v(\Box iq \rightarrow \Box\Box iq) = 1$ for any $i \in S(\neg \wedge \Box)$, and so $v(\Box(\Box q \rightarrow \Box\Box q) = 1$.

Now let $v$ be any $S(\neg\wedge)$-valuation, take any $i \in S(\neg\wedge)$, and let $A := ip$. By definition $A$ is non-modal. Suppose for reductio that $v(\Diamond A \wedge \Diamond\neg A \wedge \Box(A \rightarrow \Box A)) = 1$. Since $v(\Diamond A) = 1$ and $A$ is non-modal, there is some valuation of propositional logic, $u$, on the non-modal language that makes $A$ true. Now construct a substitution $j$ that maps each letter $p$ in $A$ to a literal that is true in $v$, if $u(p) = 1$, and a literal that is false in $v$ if $u(p) = 0$, and in such a way that we never pick two literals containing the same letter. Since for each letter $p_m$ occurring in $A$ we have $v(jp_m) = u(p_m)$, we have that $v(A(jp_1 \ldots jp_n)) = u(A(p_1 \ldots p_n)$ (writing $A$ as a function of its sentence letters), which means $v(jA) = u(A) = 1$. Moreover, since $v(\Diamond\neg A) = 1$, $A$ is not a tautology, and so $jA$ is not either (from the way it was constructed). So $v(\Box jA) = 0$ (since if $B$ is non-modal, $v(\Box B) = 1$ only if $B$ is a tautology). So we have $v(jA) = 1$ and $v(\Box jA) = 0$ so $v(jA \rightarrow \Box jA) = 0$ and so $v(\Box(A \rightarrow \Box A)) = 0$. This contradicts the assumption that $v(\Diamond A \wedge \Diamond\neg A \wedge \Box(A \rightarrow \Box A)) = 1$.

Thus $v(\Diamond ip \wedge \Diamond\neg ip \wedge \Box(ip \rightarrow \Box ip)) = 0$ for every non-modal substitution $i$; and so $v(\Diamond(\Diamond p \wedge \Diamond\neg p \wedge \Box(p \rightarrow \Box p))) = 0$ for every $S(\neg\wedge)$-valuation $v$. $\Box$

---

any rate, is that logic shouldn't say anything substantive about particular individuals, properties or relations (apart from the logical ones), so that any sentence that holds, as a matter of logic, for a particular non-logical constant should hold, as a matter of logic, for an arbitrary expression of the same type.

This proposition technically doesn't rule out the pre-validities of $S(\neg\wedge)$ and $S(\neg \wedge \Box)$ coinciding by being the inconsistent logic. We will eventually rule out this possibility by showing the existence $S(\neg\wedge)$ and $S(\neg \wedge \Box)$-valuations. It's also worth noting, however, that despite the stark contrast between the pre-validities of these two classes, for all we know the logics of these two substitution classes are identical.

### *4.2.3 Relationship to Kripke Models*

We briefly remark on the relationship between the foregoing substitutional interpretation of propositional modal logic and the more familiar interpretation due to Kripke (1959). Although the clauses for substitutional necessity are on the surface quite different from the clauses for necessity in a Kripke model, any $S$-valuation may be reconceived as a Kripke model.

**Definition 4.2.10** For any given substitution class, $S$, the *Kripke frame associated with* $S$, $(W_S, R_S)$, is defined by:

- $W_S := S$
- $R_S := \{(i, j \circ i) \mid i, j \in S\}$, that is, $R_S ik$ if and only if there is some $j \in S$ such that $k = j \circ i$.

Given an $S$-valuation, $v$, the *Kripke model associated with* $v$ is obtained by defining a valuation $V : W_S \times \mathcal{L}() \rightarrow \{0, 1\}$ on $(W_S, R_S)$ as follows[7]:

- $V(i, p_k) = v(ip_k)$.

$V$ may then be extended to arbitrary formulas in the usual manner.

In what follows we reserve the lower case letters $v$ and $u$ for $S$-valuations and use the uppercase letters $V$ and $U$ for valuations in a Kripke model.

**Proposition 4.2.11** *For any $i \in S$ and formula $A$, $V(i, A) = v(iA)$.*

***Proof*** By induction. The clauses for the sentence letters and truth functional connectives are trivial. For the modal clause, we reason as follows.

$V(i, \Box A) = 1$ if and only if, for every $j \in S$, $V(j \circ i, A) = 1$. By the inductive hypothesis, this holds iff $v(jiA) = 1$ for every $j \in S$, which holds iff $v(i\Box A) = 1$, as required. □

We will also frequently appeal to the notion of $p$-morphism of frames:

**Definition 4.2.12** (*p-morphism*) A $p$-morphism from a Kripke frame $(W, R)$ to another Kripke frame $(W', R')$ is a function $f : W \rightarrow W'$ such that:

[7] See Drake (1962).

(i) For any $x, y \in W$, if $Rxy$ then $R'f(x)f(y)$.
(ii) For any $x \in W$ and $y' \in W'$, if $R'f(x)y'$ then there exists a $y \in W$ such that $Rxy$ and $f(y) = y'$.

If $f$ is a $p$-morphism between the frames of the models $(W, R, V)$ and $(W', R', V')$ and $f$ preserves the truth values of some sentences, $B_1 \ldots B_n$,—i.e., $V(x, B_k) = V'(f(x), B_k)$ for $k = 1 \ldots n$ and any $x \in W$—then it may be shown that for any sentence $A$ constructed from $B_1 \ldots B_n$ using only the truth-functional and modal connectives, $V(x, A) = V'(f(x), A)$. If $f$ preserves the truth values of all the letters (so, given the above, preserves the truth values of all sentences), then $f$ is sometimes called a $p$-morphism between the models $(W, R, V)$ and $(W', R', V')$. Given the above correspondence between valuations and Kripke models, we can talk about $p$-morphisms between substitution classes and Kripke frames.

**Definition 4.2.13** (*p-morphism between substitution classes and Kripke frames*) A $p$-morphism from a substitution class $S$ to a Kripke frame $(W', R')$ is a $p$-morphism from the frame $(W_S, R_S)$ to $(W', R')$.

## 4.3 The Existence and Uniqueness of *S*-Valuations

If there were no valuations for a given substitution class, $S$, the concepts of prevalidity and validity would be trivial and uninteresting. However, given a substitution class $S$, the existence of an $S$-valuation is not always obvious. Consider, for instance, a valuation for the full substitution class mapping letters to arbitrary sentences of $\mathcal{L}$. While the truth values of conjunctions and negations are determined by the truth values of sentences of lower complexity (the conjuncts or the negatum), the truth value of modal formulas is not. The truth value of $\Box A$ is determined by the truth values of $iA$ for all possible substitutions of letters within $A$, including substitutions of higher complexity: $i$ might map some of the letters appearing in $A$ to $A$ itself for example, yielding a potentially vicious circularity.[8]

In many cases the circularity is not vicious. Let's say that a substitution $i$ is non-modal iff $ip_k$ is a formula of the propositional calculus for each $k$. The following proposition shows that whenever $S$ is a class of non-modal substitutions, $S$-valuations exist and can be constructed in a familiar inductive manner. Indeed, any assignment to the sentence letters extends to a unique $S$-valuation on $\mathcal{L}$.

**Proposition 4.3.1** *Suppose that $S$ is a class of non-modal substitutions and that $v^- : \mathcal{L}() \to \{0, 1\}$ is a truth-value assignment on the sentence letters of $\mathcal{L}$. Then there exists a unique $S$-valuation extending $v^-$.*

***Proof*** $v$ may be defined inductively on the modal degree of the the formula. In particular, suppose that $v(A)$ has been defined for every formula of modal degree $n$,

[8] See the related remarks in Kripke (1976, p. 332) concerning the substitutional interpretation of the quantifiers.

and we are attempting to evaluate $v(\Box A)$ and $v(\Diamond A)$ of modal degree $n+1$. Since each $i \in S$ is non-modal, $iA$ will have modal degree $n$, and so $v(\Box A)$ and $v(\Diamond A)$ may and must be defined as $\min_{i \in S} v(iA)$ and $\max_{i \in S} v(iA)$, respectively. □

When $S$ contains substitutions with modal formulas in their range, the constraints of Definition 4.2.6 can no longer be met through a straightforward inductive construction. Nonetheless, we might conjecture that an analogue of Proposition 4.3.1 holds for the full substitution class:

**Conjecture 4.3.2** (*The Uniqueness Conjecture*) Suppose that $S$ is the class of all substitutions on $\mathcal{L}$, and suppose that $v^- : \mathcal{L}() \to \{0, 1\}$ is a truth-value assignment on the sentence letters. Then there exists a unique $S$-valuation $v$ extending $v^-$.

Indeed this conjecture is a slight variant of a conjecture made by Friedman (1975).[9] The existence portion of the conjecture was settled independently by Kit Fine and Tadeusz Prucnal.[10] Prucnal's method is rather indirect, and goes via a solution to a related problem concerning intuitionistic logic. In Sect. 4.3.1 we present Fine's direct proof of the existence of a valuation of the full substitution class. The uniqueness part of the conjecture is still open.

For some classes of modal substitutions we can prove the existence of an $S$-valuation through a slightly different inductive construction. To illustrate, consider the 'positive substitutions' from Example 4.2.4, that map formulas to positive formulas (in which letters only appear under an even number of negations). Let $P$ be the class of all positive substitutions on $\mathcal{L}$.

**Proposition 4.3.3** *There exist $P$-valuations.*

***Proof*** Let $X_0$ be the set of positive formulas, and $X_{n+1}$ the truth functional combinations of formulas from $\{A, \Box A \mid A \in X_n\}$. We define a valuation inductively as follows. Each formula of $X_0$ is treated as true. For any positive substitution $i$ and sentence $A \in X_n$, it is readily shown that $iA$ is also in $X_n$. Thus assuming the valuation is defined on $X_n$ we can extend the valuation (uniquely) to formulas of the form $\Box A$ for $A \in X_n$ and thus to arbitrary formulas in $X_{n+1}$. □

### 4.3.1 The Existence of Valuations of the Full Substitution Class

Next we settle the harder question of whether there are any valuations for the full substitution class. We will show that every truth-value assignment to the sentence letters, $v^- : \mathcal{L}() \to \{0, 1\}$, can be extended to a valuation of the full substitution class.

[9] Friedman's conjecture (see problem 42 of Friedman, 1975) was concerned with the existence and uniqueness of a valuation on the full substitution class making all the sentence letters true. Moreover, his version of the conjecture postulates the existence and *non*-uniqueness of such a valuation.

[10] See Prucnal (1979).

In order to do this we will introduce a special class of finite, partially ordered frames that have the property that every world can easily be characterized, in relation to the endpoints of the order, by a modal formula. Henceforth we will write $P^0(X)$ for $P(X) \setminus \emptyset$ and we will write (being somewhat sloppy) $\supseteq$ for the superset relation restricted to $P^0(X)$ when $X$ is clear from context.

**Definition 4.3.4** (*Medvedev frame*) A Medvedev frame is a frame of the form $(P^0(X), \supseteq)$ for some finite non-empty set $X$.

We write Med for the logic of Medvedev frames.

We begin with a brief informal overview of our proof strategy. The key idea is to take a countable set of disjoint models over Medvedev frames for which Med is complete, and glue them together to make a new model $M$ by placing a new root world that lies underneath the original 'component' models and that sees itself and all the worlds of the component models. The truth values of the letters at the new root world are determined by $v^-$, and at the other worlds by their truth value in the relevant component model. We then argue that a sentence $\Box A$ is true at the root of this global model if and only if every substitution instance of $A$ is true at the root. The valuation of the full substitution class is then obtained by identifying the truth value of a sentence with its truth value at the root world.

To make this precise, enumerate all of the consistent sentences of Med, $C_1, C_2, C_3, \ldots$. For each $C_i$, pick a model $M_i = (P^0(X_i), \supseteq, V_i)$ on which $C_i$ is satisfiable at some world. We may assume without loss of generality that $X_i$ and $X_j$ are disjoint when $i \neq j$. Then our global model $M = (W, \supseteq, V)$ may be defined as follows:

- $X := \bigcup_{i \in \omega} X_i$
- $W := \{w \subseteq X \mid w = X \text{ or } w \in P^0(X_i) \text{ for some } i\}$
- $V(w, p_j) = V_i(w, p_j)$ if $w \in P^0(X_i)$ for some $i$
- $V(X, p_j) = v^-(p_j)$.

Our goal is to show that for any formula $A$:

**Proposition 4.3.5** *$V(X, \Box A) = 1$ if and only $V(X, iA) = 1$ for every substitution $i$.*

The proof of the left-to-right direction is quite straightforward. The frame of $M$ is not Medvedev because $W$ is infinite, but it may nonetheless be shown to be a model of Med in the sense that $V(w, B) = 1$ for any formula $B \in$ Med and world $w \in W$ (Lemma 4.3.6 below). The left-to-right direction follows, since if $\Box A$ is true at the root $X$ of $M$, $A$ must be true in each component model, and thus must be in Med by the completeness of Med with respect to the component models. But since Med is closed under the rule of substitution, every substitution instance of $A$ is in Med, and so every substitution instance of $A$ must be true at the root world since $M$ is a model of Med.

The proof of the right-to-left direction is more involved. It suffices to show that if $\Box A$ is false at the root world $X$ then some substitution instance of $A$ is also false at $X$. If $\Box A$ is false at the root, then $A$ must be false at some world in $M$, and so

must not belong to Med since, as before, $M$ is a model of Med. So $A$ can be made false at a world in a model $N = (P^0(Y), \supseteq, U)$ over a Medvedev frame based on the set $Y = \{1, \ldots, n\}$; indeed it can be made false at the root $Y$ of such a model.[11] This model refuting $A$ encodes in a natural way the required substitution instance of $A$ that is false at the root $X$ of $M$. We now describe how to extract that substitution from the model $N$.

Without loss of generality we may assume that the terminal worlds of $N$, $\{1\}, \ldots, \{n\}$, make different propositional letters true and so there are propositional formulas $A_1, \ldots A_n$ such that $A_m$ is true at $\{m\}$ and false at $\{m'\}$ for $m' \neq m$. For reasons that will come in to play later, we will also choose $A_1 \ldots A_n$ so that they are pairwise incompatible and have a tautologous disjunction.[12] Since a singleton $\{m\}$ sees only itself, $\Box A_m$ will be true at $\{m\}$. Now a given world $Z$ (i.e., subset of $\{1, \ldots, n\}$) is uniquely determined by its singleton subsets, or equivalently the terminal worlds $\{m\}$ it sees. Thus $Z$ makes $\Diamond\Box A_m$ when $m \in Z$ and $\Diamond\Box A_m$ false when $m \notin Z$, and is the unique world of our model to do so. Thus we may define a formula, $D_Z$ that is true only at world $Z$ in $N$:

$$D_Z := \bigwedge_{m \in Z} \Diamond\Box A_m \wedge \bigwedge_{m \notin Z} \neg\Diamond\Box A_m$$

Since $N$ is finite, any set of worlds is characterized by a disjunction of these world formulas. In particular, for each propositional letter $p$ we may define a formula made only out of formulas characterizing the terminal worlds (i.e., $A_1 \ldots A_n$) that has the same truth value as $p$ at any world.

$$D_p := \bigvee_{U(Z,p)=1} D_Z$$

We can now define our desired substitution by mapping each letter $p$ to $D_p$:

$$i(p) := D_p$$

Since $p$ and $ip$ have the same truth value at every world in $N$, the result of uniformly substituting $p$ for $ip$ in any formula must have the same truth value at any world in $N$. Since $A$ is false at the root $Y = \{1, \ldots, n\}$ in $N$, $iA$ must also be false at the root $Y$.

We now show that $iA$ must also be false at the root of $M$. According to any world $w$ in our global model $M$, some of $A_1 \ldots A_n$ are possibly necessary, and others are not: so there is a *unique* world in $N$ that agrees with $w$ about which $A_1 \ldots A_n$ are possibly necessary, namely the world $\{m_1, \ldots, m_k\}$ assuming $A_{m_1}, \ldots, A_{m_k}$ are

[11] If $A$ is refuted at world $Y$ of cardinality $n$ at a non-root world of a larger Medvedev frame, the submodel generated by $Y$ is easily seen to be isomorphic to a model on the Medvedev frame $\{1, \ldots, n\}$.

[12] This may always be done, for when we are constructing a model that refutes $A$ over a Medvedev frame we are free to set the truth-values of letters not appearing in $A$ how we please. So pick sentence letters $p_1 \ldots p_k$ not appearing in $A$ such that $2^k \geq n$ and pick a surjection $\sigma : 2^k \to n$. $A_m$ may then be defined as $\bigvee_{\sigma(u)=m}(\bigwedge_{u(r)=1} p_r \wedge \bigwedge_{u(r)=0} \neg p_r)$. A valuation in which $U(\{m\}, A_m) = 1$ can be constructed by picking some representative element $u$ from $\sigma^{-1}(m)$ and letting $U(\{a_m\}, p_r) = u(r)$ for $r = 1 \ldots k$.

possibly necessary at $w$ in $M$. Note that the root of $M$, $X$, gets associated with $Y = \{1, \ldots, n\}$ since each $A_k$ is possibly necessary at the root of $M$. Let us call this function associating worlds of $M$ to worlds on $N$, $f : W \to P^0(Y)$:

$$f(w) = \{m \mid V(w, \Diamond\Box A_m) = 1\}$$

It may be shown (Lemma 4.3.7) that this function $f$ is a $p$-morphism of frames (it is here we use the fact that $A_1 \ldots A_n$ are pairwise incompatible). Now if a $p$-morphism preserves the truth values of some modal formulas $C_1, \ldots, C_k$ then it will also preserve the truth value of any formula constructed from $C_1 \ldots C_k$ using the truth-functional and modal connectives (recall Definition 4.2.12). Since by definition $f$ preserves the truth values of formulas of the form $\Diamond\Box A_k$, and the formulas $D_p$, and thus also $i A$, are constructed from these formulas using the modal and truth-functional connectives $f$ preserves the truth values of the formulas $D_p$ and thus also $i A$. Since $f$ maps $X$ to $Y$ and $i A$ is false at $Y$, $i A$ is false at the root of $M$ as required.

It remains to prove the lemmas appealed to above. The first is that $M$ is a model of Med:

**Lemma 4.3.6** *If $A \in$ Med and $w \in W$, then $V(w, A) = 1$.*

This is a consequence of a more general result—Proposition 4.6.14—that will we prove later in Sect. 4.6.2, so we will defer its proof.[13]

**Lemma 4.3.7** *$f$ is a p-morphism from the frame of $M$ to the frame of $N$.*

***Proof*** We first we show that $f$ preserves the accessibility relation. Suppose $w$ sees $v$ in $M$. Since the accessibility relation is transitive, anything possible at $v$ is possible at $w$. So $f(v) \subseteq f(w)$ and hence $f(w)$ sees $f(v)$ in $N$.

To establish the reverse condition, suppose that $f(w)$ sees a world $Z \in P^0(Y)$, so that $Z \subseteq f(w)$. We must show that there exists a $v \subseteq w$ such that $f(v) = Z$.

By definition, for each $m \in f(w)$, $w \Vdash \Diamond\Box A_m$. Thus for each $m \in f(w)$ there is a terminal world $\{a_m\} \in W$ such that $\{a_m\} \subseteq w$ and $\{a_m\} \Vdash A_m$. Let $v = \{a_m \mid m \in Z\}$. Clearly $v \subseteq w$. Moreover, we may show that $f(v) = Z$. For, given any $m \in Z$, $\{a_m\} \Vdash \Box A_m$ and $\{a_m\} \subseteq v$, so $v \Vdash \Diamond\Box A_m$, and so $m \in f(v)$. Conversely, if $v \Vdash \Diamond\Box A_m$ then $v$ sees a terminal world that makes $A_m$ true. Since $A_1 \ldots A_n$ are pairwise incompatible the only singleton $v$ sees where $A_m$ is true is $\{a_m\}$, so $a_m \in v$ and so $m \in Z$. □

We may, finally, obtain a valuation $v$ of the full substitution class by setting $v(A) = V(X, A)$.

**Theorem 4.3.8** *Every truth-value assignment $v^-$ may be extended to a valuation of the full substitution class.*

It should be observed that, in the constructed valuation, the truth of a possibility sentence $\Diamond A$, may always be witnessed by a substitution of formulas of modal degree 2.

[13] In Sect. 4.6.2 it is shown that Med is a *coherent logic* (see Definition 4.6.5), and Lemma 4.3.6 follows from the observation that the function $v$, defined by setting $v(A) := V(X, A)$, is a meta-valuation (see Definition 4.6.4).

### 4.3.2 The Uniqueness Conjecture

Proposition 4.3.1 established that for any non-modal substitution class $S$, any truth assignment to the sentence letters extends to a *unique* $S$-valuation. The analogue of Proposition 4.3.1 for the full substitution class is the principle we called Conjecture 4.3.2:

> Any truth-value assignment $v^- : \mathcal{L}() \to \{0, 1\}$ extends to a unique $S(\neg \wedge \Box)$-valuation.

Theorem 4.3.8 establishes the existence portion of the conjecture. The uniqueness conjecture is important to the study of logical necessity because it is implies that there are interpretations of $\Box$ in which $\Box$ means *valid*. Letting $S$ be the full substitution class, Conjecture 4.3.2 is equivalent to:

> For any $S$-valuation $v$, $v(\Box A) = 1$ iff $A$ is valid with respect to $S$.

Supposing the uniqueness conjecture to be true, $v(\Box A) = u(\Box A)$ for any pair of $S$-valuations, and so $v(\Box A) = 1$ iff $u(\Box A) = 1$ for all valuations $u$. Either side of the biconditional holds iff, for every valuation $u$ and substitution $i$, $u(iA) = 1$. So under the supposition, $\Box A$ will coincide with validity with respect to the full substitution class. The converse holds too. For suppose the previous claim holds. Then for any pair of $S$-valuations $u$ and $v$, $u(\Box A) = 1$ iff $A$ is valid with respect to $S$ iff $v(\Box A) = 1$; and so if $v$ and $u$ additionally agree on the sentence letters, $u = v$.

Notice that in the valuation constructed in Theorem 4.3.8, $v(\Box A) = 1$ if and only if $A \in \mathsf{Med}$. Thus an extension of a truth-value assignment, $v^-$, to a valuation for the full substitution class may be defined inductively as follows.

- $v(p) = v^-(p)$
- $v(A \wedge B) = min(A, B)$
- $v(\neg A) = 1 - v(A)$
- $v(\Box A) = 1$ iff $A \in \mathsf{Med}$

For an arbitrary valuation $v$ of $\mathcal{L}$, let $\Delta_v = \{A \mid v(\Box A) = 1\}$. Then another way to state Conjecture 4.3.2 is:

> For any valuation $v$ of the full substitution class, $\Delta_v = \mathsf{Med}$.

While we are not able to settle the uniqueness conjecture, we will here establish the weaker claim:

> For any valuation $v$ of the full substitution class, $\Delta_v \subseteq \mathsf{Med}$.

Moreover, in Sect. 4.5.2 we'll show that some principles distinctive to Med must also belong to $\Delta_v$. Thus $\Delta_v$ would appear to be tightly sandwiched between Med and a strong subsystem of Med, lending further credence to the uniqueness conjecture.

Let $v$ be a valuation for the full substitution class. For each $n$, pick a propositional partition $A_1 \ldots A_n$—i.e., $n$ propositional formulas that are pairwise inconsistent and have a tautologous disjunction. For any substitution, $i$, let $f(i) = \{m \mid v(\Diamond \Box i A_m) = 1\}$. By a $\top\bot$ substitution we shall mean a substitution whose range on $\mathcal{L}()$ is $\{\top, \bot\}$.

**Lemma 4.3.9** *For any substitution $i$ and $Y \subseteq f(i)$, there exists a substitution $j$ such that $f(j \circ i) = Y$.*

***Proof*** In this proof we will frequently appeal to the fact that if $B$ is a formula, $v$ a valuation, and $j$ a substitution such that $v(\Box jB) = 1$, then there exists a $\top\bot$ substitution $k$ such that $v(\Box kB) = 1$. For if $v(\Box jA) = 1$, for an arbitrary $j$ then the result of pre-composing $j$ with any $\top\bot$ substitution $k'$, yields $\top\bot$ substitution, $k' \circ j$, with the required property.

Suppose that $Y \subseteq f(i)$. Let $p_1 \ldots p_r$ be the letters appearing in $A_1 \ldots A_n$, and henceforth let $k$ be variable for a $\top\bot$ substitution defined only on $p_1 \ldots p_r$.

Since, by definition, $v(\Diamond\Box iA_m) = 1$ for each $m \in f(i)$, there must exist for each $m \in f(i)$ a substitution $k$ such that $v(\Box kiA_m) = 1$. And without loss of generality we may assume that $k$ is a $\top\bot$ substitution. For each $m \in f(i)$ we shall pick a representative $\top\bot$ substitution, $k_m$, such that $v(\Box kiA_m) = 1$.

Let $S$ be the set of all $\top\bot$ substitutions defined on $p_1 \ldots p_r$, and let $S' = \{k_m \mid m \in Y\}$. Our goal will be to find a substitution $j$ that 'filters' the substitutions in $S$ leaving all and only substitutions in $S'$ behind: i.e., for any $k \in S$, $k \circ j$ is equivalent to a substitution in $S'$ and every substitution in $S'$ is equivalent to something of the form $k \circ j$ for some $k \in S$. Here two substitutions, $k$ and $k'$ are said to be equivalent when $v(\Box(kp \leftrightarrow k'p)) = 1$ for any sentence letter $p$ from $p_1 \ldots p_r$.

Because $A_1 \ldots A_n$ are pairwise inconsistent and have a tautologous disjunction, we also know that for each $\top\bot$ substitution $k$ there is a unique sentence $A_m$ such that $v(\Box kiA_m) = 1$. For a given $k$ we will call this sentence $A^k$. Now pick any surjection $\sigma : S \to S'$ and define a substitution $j$ as follows:

$$j(p) = \bigvee_{\sigma(k)(p)=\top} A^k$$

for each $p \in \{p_1 \ldots p_r\}$

We now show that for any $\top\bot$ substitution $k$ on $p_1 \ldots p_n$, $\sigma(k)$ and $k \circ j$ are equivalent substitutions. We will show for each letter $p$, $v(\Box kjp) = 1$, if and only if $\sigma(k)(p) = \top$. Suppose that $\sigma(k)(p) = \top$. So $A^k$ is a disjunct of $\bigvee_{\sigma(k)(p)=\top}(A^k) = j(p)$, and so $k(A^k)$—which recall is necessary in $v$ by definition—is a disjunct $kj(p)$. Conversely, suppose that $v(\Box kjp) = 1$. Since the $A^k$ are pairwise incompatible, this can only happen if $A^k$ is one of the disjuncts of $j(p)$, and thus can only happen if $\sigma(k)(p) = \top$.

This delivers us our desired result: for if $m \in Y$ then, because $\sigma$ is surjective, there is some $\top\bot$ substitution such that $\sigma(k) = k_m$, and $k_m$ and $k \circ j$ are equivalent. Since $v(\Box k_m iA_m) = 1$, by definition, and $k_m$ is equivalent to $k \circ j$, $v(\Box kjiA_m) = 1$, and thus $v(\Diamond\Box jiA_m) = 1$. So $m \in f(j \circ i)$. Conversely if $m \in f(j \circ i)$, this means $v(\Diamond\Box jiA_m) = 1$ and so there is substitution $k$ (which we may assume without loss of generality to be $\top\bot$) such that $v(\Box kjiA_m)$. $k \circ j$ is equivalent to $\sigma(k)$ which may be seen to equivalent to $k_m$. Since $v(\Box k_m iA_m) = 1$, $m \in Y$. □

The significance of Lemma 4.3.9 is this. Let $(W, R, V)$ be the Kripke model associated with $v$ (so $W$ is the class of substitutions, $R = \{(i, j \circ i) \mid i, j \in W\}$ and $V(i, p) = v(ip)$). Then:

**Corollary 4.3.10** *$f : W \to P^0(X)$ is a p-morphism of frames from $(W, R)$ to $(P^0(X), \supseteq)$ where $X = \{1, \ldots, n\}$.*

***Proof*** Since $Rik$ iff there exists a $j$ such that $k = j \circ i$, the back-and-forth condition amounts to the claim that if $f(i) \supseteq Y$ then there exists a $j$ such that $f(j \circ i) = Y$. This is just Lemma 4.3.9.

It remains to show that $f(j \circ i) \subseteq f(i)$ for every $i, j \in W$. If $m \in f(j \circ i)$ then $v(\Diamond\Box jiA_m) = 1$ and so there is a substitution $k$ such that $v(\Box kjiA_m) = 1$. In which case the substitution $k \circ j$ witnesses the truth of $v(\Diamond\Box iA_m) = 1$ so $m \in f(i)$. □

Let $X = \{1, \ldots, n\}$. For a given $Y \subseteq X$ we define, as before:

$$D_Y := \bigwedge_{m \in Y} \Diamond\Box A_m \wedge \bigwedge_{m \notin Y} \neg\Diamond\Box A_m$$

As before, $D_Y$ and $D_Z$ are inconsistent when $Y \neq Z$. Observe, also, that for any substitution $i$, $v(i D_{f(i)}) = 1$ since, by definition, $f(i)$ is just the set of $m$ such that $v(\Diamond\Box iA_m) = 1$.

Suppose that $(P^0(X), \supseteq, V)$ is a Medvedev model, that $X = \{1, \ldots, n\}$ and that $A_1 \ldots A_n$ are a partition of propositional formulas such that $V(\{m\}, A_m) = 1$. We may define a substitution as follows:

$$i(p) = \bigvee_{V(Y,p)=1} D_Y$$

**Proposition 4.3.11** *For any formula $A$ and substitution $j$, $v(jiA) = V(f(j), A)$.*

***Proof*** We may prove this by induction of formula complexity.

Base case: Suppose that $V(f(j), p) = 1$. So $D_{f(j)}$ is a disjunct of $i(p)$, and since $j$ distributes over disjunctions, $j D_{f(j)}$ is a disjunct of $jip$. By our observation $v(j D_{f(j)}) = 1$ so $v(jip) = 1$.

Conversely, suppose $v(jip) = 1$. So $v(j D_Y) = 1$ for some $Y$ such that $V(Y, p) = 1$. By our observation we know that $v(j D_{f(j)}) = 1$. Since the $D_Z$'s are pairwise incompatible (and, thus, so are their substitution instances) it follows that $v(j D_Y) = v(j D_{f(j)}) = 1$ only if $f(j) = Y$. So $V(f(j), p) = V(Y, p) = 1$.

Inductive step: the negation and conjunction cases are straightforward, and the $\Box$ case follows from the fact that $f$ is a p-morphism. Explicitly: if $v(\Box jiA) = 0$ then $v(kjiA) = 0$ for some substitution $k$. So $V(f(kj), A) = 0$ by the inductive hypothesis.

Conversely, if $V(f(j), \Box A) = 0$ then $V(Y, A) = 0$ for some non-empty $Y \subseteq f(j)$. By Lemma 4.3.9 there exists a $k$ such that $f(k \circ j) = Y$, and so by the inductive hypothesis, $v(kjiA) = V(f(k \circ j), A) = V(Y, A) = 0$. □

**Theorem 4.3.12** *If $v$ is a valuation of the full substitution class, $\Delta_v \subseteq$ Med.*

***Proof*** It suffices to show that any sentence $C$ consistent in Med is consistent in $\Delta_v$: there is some substitution $k$ such that $v(kC) = 1$. If $C$ is consistent in Med then there is a Medvedev model $(P^0(X), \supseteq, V)$ such that $C$ is true at some world $Y \subseteq X$. Construct a substitution $i$ as above, and pick some substitution $j$ such that $f(j) = Y$ (this is possible by Lemma 4.3.9). Then $v(jiC) = V(f(j), C) = V(Y, C) = 1$. $j \circ i$ is the required substitution. □

## 4.4 The Logic of Logical Necessity

Which principles of modal logic are valid on the logical interpretation of □? The existence of an identity substitution immediately ensures the validity of the the T axiom, $\Box A \to A$. For given any substitution class $S$, and $S$-valuation, $v$, it follows that $v(\Box A) = 1$ only if $v(iA) = 1$ for any substitution and, in particular, that $v(\iota A) = v(A) = 1$ when $\iota$ is the identity substitution. Similarly, the closure of the substitutions under composition ensures the validity of the S4 axiom, since if $v(\Box A) = 1$ then, for any pair of substitution $i, j \in S$, their composition $i \circ j$ is also in $S$, and so $v((i \circ j)A) = 1$. For fixed $j$, if $v(ijA) = 1$ for every $i \in S$, it follows that $v(\Box jA) = 1$; and since this holds for every $j \in S$, $v(\Box\Box A) = 1$. Indeed, McKinsey showed that for every substitution class $S$, $L(S)$ contains the theorems of S4.[14]

**Theorem 4.4.1** (McKinsey) *Let $S$ be any substitution class.*

1. *The set of validities $L(S)$ is closed under the rules of necessitation, modus ponens and uniform substitution.*
2. *The formulas $\Box A \to \Box\Box A$, $\Box A \to A$, $\Box(A \to B) \to (\Box A \to \Box B)$ are all valid.*

*Thus every theorem of* S4 *is valid with respect to the class of $S$-valuations.*

***Proof*** This result appears in McKinsey (1945). We may obtain it directly from Proposition 4.2.11 by noting that the Kripke model associated with any $S$-valuation $v$ is transitive and reflexive, as ensured, respectively, by the Composition and Identity conditions on substitution classes. It follows that for any $A$ and $B$, $\Box A \to \Box\Box A$, $\Box A \to A$ and $\Box(A \to B) \to \Box A \to \Box B$ are true in any valuation. Closure under the rules of necessitation, modus ponens and uniform substitution follows straightforwardly from the definitions of validity and $S$-valuation. □

[14] Drake (1962) improves on this result by showing that the intersection of the logics $L(S)$ for every substitution class $S$ is exactly S4.

### 4.4.1 The Brouwer and McKinsey Axioms

Which further principles of modal logic might be valid under the logical interpretation of necessity? Taking the logic of metaphysical necessity as our cue, one might wonder if the Brouwerian axiom

B $A \rightarrow \Box\Diamond A$

Is valid. The result of adding B to S4 yields S5, a logic commonly supposed to be the logic of metaphysical necessity. In correspondence, McKinsey offers the following argument against the truth of the Brouwerian axiom.[15] While is certainly true that sugar is sweet, and vinegar is not, the Brouwerian axiom, if it were true, would further imply that it's necessarily possible that sugar is sweet and vinegar is not sweet. But given the left-to-right direction of the substitutional constraint, substituting sugar for vinegar, we may infer the absurd conclusion that it's possible that sugar is sweet and sugar is not sweet.

This conclusion is not peculiar to McKinsey's definition of logical necessity either, since it is also a consequence of our more general class of metalogical constraints on logical necessity. For as we noted there, so long as our account of logical truth, $\Delta$, is closed under the rule of substitution we have the left-to-right direction of McKinsey's substitutional constraint. (The reader may have noticed that we actually *do* have a candidate notion of logical truth which isn't closed under the rule of substitution—pre-validity. We'll investigate this interpretation in Sect. 4.4.2, where we will see that it is indeed consistent with the logic of S5.)

What is special about the substitution of 'sugar' for 'vinegar' in 'sugar is sweet and vinegar is not' is that no further substitutions to the resulting sentence can change its truth value. Let us sharpen this notion:

**Definition 4.4.2** A substitution $i \in S$ is terminal with respect to $S$ if and only if $v(\Box i\, A) = 1$ or $v(\Box \neg i\, A) = 1$ for every sentence $A$ of $\mathcal{L}$ and $S$-valuation $v$. A terminal substitution, $i$, *matches* to a $\top\bot$-substitution $j$ if and only if $v(\Box ip) = 1$ when $jp = \top$ and $v(\Box \neg ip) = 1$ when $jp = \bot$.

Finally, say that a substitution class contains all terminal substitutions when it has a terminal substitution matching any $\top\bot$ substitution, where, as before a $\top\bot$-substitution is a substitution whose range on $\mathcal{L}()$ is $\{\top, \bot\}$.

Clearly any substitution class containing a $\top\bot$ substitution, such as $S(\top\bot\ldots)$ or $H$ (the class of substitutions that leave letters alone or replace them with $\top$ or $\bot$) has a terminal substitution. However, many other formulas can play the role of $\top$, such as $(p_0 \vee \neg p_0)$ or $\Box p_0 \rightarrow \Box\Box p_0$, and similarly many formulas can play the role of $\bot$. So one need not have $\top$ and $\bot$ in the range of substitutions in order to have terminal substitutions.

[15] See Anderson and Belnap (1975, pp. 122–123).

The presence of terminal substitutions imposes its own distinctive modal logic. The McKinsey axiom says:

M $\Box\Diamond A \rightarrow \Diamond\Box A$

It characterises the Kripke frames in which each world sees a terminal world: a world that sees only itself.[16]

**Proposition 4.4.3** (McKinsey) *Let $S$ be a substitution class containing a terminal substitution. Then the theorems of* S4M *are valid over the class of $S$-valuations.*

***Proof*** Since, by Theorem 4.4.1, the validities of any substitution class are closed under modus ponens and necessitation and contain the axioms of S4, it remains only to show that the McKinsey axiom M is valid. If $\Box\Diamond A$ is true in some arbitrary $S$-valuation $v$, then for some terminal substitution $i \in S$, $v(\Diamond i A) = 1$. Since $i$ is terminal, $v(\Box i A) = 1$, and so $v(\Diamond\Box A) = 1$. □

The McKinsey axiom is compatible with Brouwer's axiom. However, the only consistent modal logic containing S4, McKinsey's axiom, and Brouwer's axiom is the trivial modal logic Triv in which $\Box$ just means *true*.[17] The modal logic Triv has as its characteristic axiom:

Triv $\Box A \leftrightarrow A$

Triv is the logic of the trivial substitution class that has the identity substitution $\iota$ as its sole element.

Indeed, according to the Brouwerian axiom, truths of the form $\Box\Diamond A$ are quite commonplace: for any sentence $A$, either $\Box\Diamond A$ or $\Box\Diamond\neg A$ is true, depending on whether $A$ or $\neg A$ is true. By contrast, in a substitution class containing all terminal substitutions, truths of the form $\Box\Diamond A$ are quite special: it may be shown that the sentence $\Box\Diamond A$ is true *only if* $A$ is a theorem of Triv, or equivalently, if the result of deleting all modal operators from $A$ is a tautology. Let us write $A^-$ for the propositional formula that results from deleting all modal operators from $A$. Then:

**Proposition 4.4.4** *Suppose $S$ is a substitution class for which there is at least one $S$-valuation and which contains all terminal substitutions. Then the following are equivalent.*

---

[16] McKinsey actually considers the formula:

F $\Box\Diamond A \wedge \Box\Diamond B \rightarrow \Diamond(A \wedge B)$

Which is only equivalent to M in the context of S4. The principle M appears to originate from Cresswell and Hughes (1996).

[17]

1. $A \rightarrow \Box\Diamond A$ (an instance of B)
2. $\Box\Diamond A \rightarrow \Diamond\Box A$ (an instance of M)
3. $\Diamond\Box A \rightarrow \Box A$ (an equivalent of the 5 axiom)
4. $A \rightarrow \Box A$ from 1 to 3
5. $\Box A \rightarrow A$ by T.

1. $\Box\Diamond A$ *is true in all S-valuations.*
2. $\Box\Diamond A$ *is true in some S-valuation.*
3. $A^-$ *is a tautology.*
4. $A$ *is a theorem of* Triv.

***Proof*** The equivalence of 3 and 4 is straightforward. Clearly 1 implies 2. We show that 2 implies 3 and 3 implies 1, establishing the equivalence of 1–3.

We may show by a straightforward induction that if $i$ is a terminal substitution, then $v(iB) = v(i(B^-))$ for any $B$. The only non-trivial case is to show that $v(i\Box C) = v(i(\Box C)^-)$. $v(i\Box C) = v(\Box iC) = 1$ iff $v(iC) = 1$ since $i$ is terminal. Moreover, by the inductive hypothesis, $v(iC) = 1$ iff $v(i(C^-)) = 1$. But $v(i(\Box C)^-) = v(i(C^-))$. So $v(i\Box C) = v(i((\Box C)^-))$.

Assume 2, and let $v$ be an $S$-valuation such that $v(\Box\Diamond A) = 1$. Thus for every terminal substitution $i$, $v(\Diamond iA) = v(iA) = 1$. So $v(iA^-) = 1$ for every terminal substitution $i$, which clearly implies that $A^-$ is a tautology. This establishes 3.

Now suppose that $A^-$ is a tautology and $j$ an arbitrary $S$ substitution. Let $i$ be any terminal substitution and $v$ any $S$-valuation. $v(ijA) = v(i((jA)^-))$ from before. But if $A^-$ is a tautology, then $(jA)^-$ is a substitution instance of $A^-$ of the letters by sentences of propositional logic (specifically the substitution $j^-$ where $j^-(p_k) = (jp_k)^-$). So $(jA)^-$ is a tautology too, and hence $v(ijA) = v(i((jA)^-) = 1$. Thus $i$ witnesses the truth of $v(\Diamond jA) = 1$; and since $j$ was an arbitrary $S$ substitution, $v(\Box\Diamond A) = 1$. □

### *4.4.2 Carnap's Theory of Logical Necessity*

Earlier we reported a general argument, due to McKinsey, that the logic of logical necessity does not contain the Brouwerian axiom. It rested on the proposed constraint connecting logical necessity and logical truth, along with the assumption that logical truth is closed under the rule of substitution. Might we resist this argument by adopting a conception of logic that is not closed under the rule of substitution? Interestingly, an early example of a theory of logical necessity in Carnap (1946, 1947, pp. 173–177) results in a notion of logical truth that fails to be closed under the rule of substitution (see Cresswell, 2013), whilst simultaneously validating all the theorems of S5, including all instances of Brouwer's axiom.

Indeed, we have encountered our own candidate notion of logical truth not closed under substitution: pre-validity. This raises the question of whether there might be a substitutional interpretation of $\Box$ as pre-validity:

> Is there a substitution class $S$ such that, for any $S$-valuation $v$, $v(\Box A) = 1$ if and only if $A$ is pre-valid with respect to $S$?

Indeed, we will show that there is and that the pre-validities of this class coincide with Carnap's notion of logical truth for propositional modal logic.[18]

Recall from Example 4.2.5 in Sect. 4.2.1 that $K$ is the class of substitutions $i$ such that $i(p_k)$ is either $p_k$ itself, or the result of prefixing a finite string of negations in front of $p_k$. Note that there is in effect no difference between substitutions with the same parity: if for each $k$, $ip_k$ and $jp_k$ agree on whether they contain an even or odd number of negations, then for the purposes of valuation we might as well treat these substitutions as the same. However this redundancy is necessary to meet the formal requirement that a substitution class be closed under substitution.

**Proposition 4.4.5** *For any $K$-valuation $v$, $v(\Box A) = 1$ iff $A$ is pre-valid with respect to $K$.*

***Proof*** Given any $K$-valuation $v$ and substitution $i \in K$, let $(iv)$ be the unique $K$-valuation defined by setting $(iv)(p_k) = v(ip_k)$ (so $(iv)(p_k) = 1$ iff $ip_k$ has an even number of negations and $v(p_k) = 1$, or $ip_k$ has an odd number of negations and $v(p_k) = 0$).

We show by induction on the modal degree of $A$ that for any $v$ and $i \in K$:

$$v(iA) = (iv)(A)$$

The identity holds of the sentence letters by construction and is clearly inherited over truth-functional compounds. So suppose that for sentences of modal degree $n$, $v(iA) = (iv)(A)$ for any $K$-valuation $v$ and substitution $i \in S$. Recall that $v(\Box iA) = 1$ iff $v(jiA) = 1$ for every substitution $j \in K$. Now for any $j$, we may apply the inductive hypothesis to $v$ and the substitution $j \circ i$ to get $v(jiA) = (jiv)(A)$; and by similarly applying the inductive hypothesis to $iv$ and $j$, we get that $(jiv)(A) = (iv)(jA)$. So $v(jiA) = 1$ for every $j \in S$ iff $(iv)(jA) = 1$ for every $j \in S$, which holds iff $(iv)(\Box A) = 1$, as required.

We may now prove the proposition. Note that for a fixed $v$, every $K$-valuation $u$ is of the form $iv$ for some substitution $i \in K$. (Set $i(p_k) = p_k$ if $u(p_k) = v(p_k)$, and let $i(p_k) = \neg p_k$ if $u(p_k) \neq v(p_k)$.) So $v(\Box A) = 1$ iff $v(iA) = 1$ for every $i \in K$, iff $(iv)(A) = 1$ for every $i \in K$, iff $u(A) = 1$ for every $K$-valuation, iff $A$ is pre-valid with respect to $K$. □

It is also easy to verify that all the theorems of S5 are valid with respect the class of $K$ valuations.

We now compare the pre-logic of our substitution class with the propositional modal logic of Carnap's theory of logical necessity.[19] Formulas of propositional

[18] See Carnap (1946). The propositional fragment of Carnap's logic has been treated in Cresswell (2013), Thomason (1973), Hendry and Pokriefka (1985). A related project is carried out in Cocchiarella (1974); see also Carroll (1978).

[19] A presentation of the propositional fragment of Carnap's theory may be found in Cresswell (2013).

modal logic are evaluated with respect to truth-value assignments to sentence letters, $v^- : \mathcal{L}() \to \{0, 1\}$. We define what it means for a sentence $A$ of propositional *modal* logic, $\mathcal{L}(\wedge\neg\Box)$, to be true at a truth-value assignment $v^-$, written $v^- \models A$, as follows:

- $v^- \models p_k$ iff $v^-(p_k) = 1$
- $v^- \models A \wedge B$ iff $v^- \models A$ and $v^- \models B$
- $v^- \models \neg A$ iff $v^- \not\models A$
- $v^- \models \Box A$ iff $u^- \models A$ for every truth-value assignment $u^-$.

Thus $\Box A$ can be taken to mean in an extended sense that $A$ is a tautology. A sentence $A$ is *C-valid* iff $v^- \models A$ for every truth-value assignment $v^-$. $C$-validity is not closed under the rule of substitution: for instance, $\neg\Box p_0$ is $C$-valid, while $\neg\Box\neg(p_0 \wedge \neg p_0)$ is not. Much like our Proposition 4.4.5, a sentence of the form $\Box A$ is true at a propositional valuation iff $A$ is $C$-valid. Indeed, we can show:

**Proposition 4.4.6** *$A$ is $C$-valid iff $A$ is pre-valid with respect to $K$.*

***Proof*** Given a $K$-valuation $v$, we write $v^-$ for corresponding the truth-value assignment: $v \restriction_{\mathcal{L}()}$. By Proposition 4.3.1 every truth-value assignment is identical to $v^-$ for some unique $K$-valuation $v$, and so we lose no generality by restricting attention to truth-value assignments of the form $v^-$. We shall show by induction that for every $K$-valuation $v$, $v(A) = 1$ iff $v^- \models A$.

The identity clearly holds for sentence letters. The clauses for the truth functional connectives are straightforward. $v^- \models \Box A$ iff $u^- \models A$ for every truth-value assignment $u^-$. So by the inductive hypothesis, $u(A) = 1$ for every $K$-valuation $u$. Now for every substitution $i \in K$, $iv$ (as defined in Proposition 4.4.5) is a $K$-valuation, so $(iv)(A) = 1$ and since $(iv)(A) = v(iA)$, $v(iA) = 1$ for every $i \in K$. Thus $v(\Box A) = 1$. Conversely, since every $K$-valuation is of the form $iv$ for some $i \in K$, if $v(\Box A) = 1$, then $u(A) = 1$ for every valuation, so by the inductive hypothesis $u^- \models A$ for every $u$ and so $v^- \models \Box A$.

As we saw, $A$ is pre-valid iff $v(\Box A) = 1$. By the above, $v(\Box A) = 1$ iff $v^- \models \Box A$ iff $A$ is $C$-valid. This completes the proof. □

## 4.5 The Logic of Specific Substitution Classes

The foregoing remarks give us some idea of what sort of logical principles must be a part of the logic of logical necessity. The situation becomes more intricate when we look at specific substitution classes. We shall see that the logics of some, though not all, substitution classes include the Grzegorczyk axiom Grz

Grz $\quad \Box(\Box(A \to \Box A) \to A) \to A$

A principle that is not a theorem of S4M.

We shall also see that the logics of some, though not all, substitution classes include an axiom we call "the subset principle" that is distinctive of Med.

### 4.5.1 The Logic of Non-modal Substitution Classes

In this section we investigate various classes of of non-modal substitutions—substitutions that map letters to non-modal propositional formulas. In his book *Philosophical Applications of Modal Logic*, Lloyd Humberstone provides an interpretation of McKinsey's theory of modality in terms of (what we have called) the class of Humberstone substitutions, $H$. These are substitutions that either map a letter to itself, $\top$ or $\bot$. After noting that every theorem of S4M is valid with respect to this substitution class (Proposition 4.4.1), he poses the following questions (p. 168):

1. Are the theorems of S4M exactly the validities with respect to the class $H$?
2. How sensitive is the logic to the exact substitution class used? Is the logic of $H$ the same as the logic of $S(\top\bot)$ or the logic of the full substitution class?

The first question may be answered negatively. In particular, the Grzegorczyk axiom Grz above is valid with respect to the class of $H$-valuations but is not a theorem of S4M.[20]

**Proposition 4.5.1** Grz *is valid with respect to* $H$*-valuations.*

***Proof*** Let $v$ be an arbitrary $H$-valuation and $A$ an arbitrary formula. We suppose that $v(A) = 0$ and then show that $v(\Box(\Box(A \to \Box A) \to A)) = 0$. Our goal is thus to find a substitution $i$ such that $v(iA) = 0$ and $v(\Box(iA \to \Box iA)) = 1$.

Without loss of generality, we may restrict our attention to substitutions that map sentence letters not appearing in $A$ to $\top$. Partially order these remaining substitutions as follows: $i \leq j$ when there exists a $k$ such that $j = k \circ i$ (or, equivalently, for all $n$, $j(p_n) = \top$ whenever $i(p_n) = \top$ and $j(p_n) = \bot$ whenever $i(p_n) = \bot$).

Because there are only finitely many letters in $A$, there are only finitely many substitutions in this ordering. Since $v(A) = 0$, there must be a maximal substitution, $i$, with respect to this ordering such that $v(iA) = 0$. This means that $v(iA) = 0$ and that for any $j > i$ $v(jA) = 1$. We now show that $v(\Box(iA \to \Box iA) = 1$. For any $j$, either (a) $j \circ i = i$, so $v(jiA) = 0$ or else (b) $j \circ i > i$, in which case $v(\Box jiA) = 1$, for $v(kjiA) = 1$ for any $k \in H$, since $i$ was the maximal substitution making $A$ false and $k \circ j \circ i \geq j \circ i > i$. Either way, $v(jiA \to \Box jiA) = 1$ and since $j$ was arbitrary, $v(\Box(iA \to \Box iA) = 1$. □

We can also use Grz to make progress with Humberstone's second question regarding $S(\top\bot)$. Recall that the difference between $S(\top\bot)$ and $H$ is that while $H$ substitutions must map $p$ to itself or to $\top$ or $\bot$, an $S(\top\bot)$ substitution can map $p$ to any other letter, in addition to $\top$ and $\bot$. We can now show:

**Proposition 4.5.2** Grz $\notin L(S(\top\bot))$ *and, specifically,* Grz *may be invalidated through the instance* $A = p \to \Box p$.

[20] To see that Grz is not a theorem of S4M, consider the frame $(\{0, 1, 2\}, \{(0, 1), (1, 0), (0, 2), (1, 2)\} \cup \{(w, w) \mid w = 0, 1, 2\})$ and the model over it in which $p$ is true at 0 and 2, but not at 1. Since it is transitive, reflexive and every world sees a terminal world, the theorems of S4M hold in this model, but $\Box(\Box(p \to \Box p) \to p)$ holds at 1 while $p$ does not.

***Proof*** Let $v$ be a $S(\top\bot)$-valuation with a true and a false letter, $p$ and $q$. We will show that $v(\Box(\Box(A \to \Box A) \to A)) = 1$ and $v(A) = 0$ when $A = p \to \Box p$. Let $i$ be any substitution in $S(\top\bot)$, and suppose $v(i(p \to \Box p)) = 0$. It straightforwardly follows that $ip$ cannot be $\top$ or $\bot$, and so it must be some sentence letter $r$. Let $j$ be a substitution such that $jr = q$, the false sentence letter.

Then $v(jiA) = v(q \to \Box q) = 1$ since $q$ is false, and $v(ji\Box A) = v(\Box(q \to \Box q)) = 0$ since $p \to \Box p$ is a substitution instance of $q \to \Box q$ and is false. Thus $v(jiA \to \Box jiA) = 0$, and so $v(\Box(iA \to \Box iA)) = 0$. So we have shown that, for an arbitrary substitution $i \in S(\top\bot)$, when $v(iA) = 0$ then $v(\Box(iA \to \Box iA)) = 0$ and hence that $v(\Box(\Box(A \to \Box A) \to A)) = 1$. Finally, $v(p \to \Box p) = 0$ since $p$ is true and $\Box p$ false in $v$. □

We can extend the above line of argument to show that Grz is invalid with respect to the class of all non-modal substitutions $S(\neg\wedge)$. In order to do this, we need to appeal to the following fact about propositional logic:

**Lemma 4.5.3** *Suppose that $A$ is a sentence of the propositional calculus that is neither tautologous nor contradictory, and that $B$ is any other sentence (possibly involving modal operators). Then there exists a substitution $i \in S(\neg \wedge \Box)$ such that $iA$ is equivalent to $B$.*

*When $B \in \mathcal{L}(\neg\wedge)$, this substitution will be in $S(\neg\wedge)$. When $B \in \mathcal{L}(\top\bot\neg)$, there will similarly exist a substitution $i \in S(\top\bot\neg)$ such that $iA$ is equivalent to $B$.*

***Proof*** Let $p_1 \ldots p_n$ be the sentence letters in $A$ and consider the truth-value assignments to $p_1 \ldots p_n$. Since $A$ is neither tautologous nor contradictory, there are assignments $v$ and $v'$ making $A$ true and false respectively. Since $v'$ may be obtained from $v$ by picking a letter $p_k$ from $p_1 \ldots p_n$ and flipping its truth value, and repeating this as many times as necessary, there must exist some assignment $u$ and letter $p_k$ such that $u$ makes $A$ true, but the assignment that results from flipping the truth value of $p_k$ in $u$ makes $A$ false. We can then define the desired substitution as follows:

$$ip_m = \begin{cases} B & \text{if } m = k \text{ and } u(p_k) = 1 \\ \neg B & \text{if } m = k \text{ and } u(p_k) = 0 \\ \neg(p_0 \wedge \neg p_0) & \text{if } m \neq k \text{ and } u(p_k) = 1 \\ p_0 \wedge \neg p_0 & \text{if } m \neq k \text{ and } u(p_k) = 0 \end{cases}$$

Clearly $i \in S(\neg\wedge)$ when $B \in \mathcal{L}(\neg\wedge)$.

For $B \in \mathcal{L}(\top\bot\neg)$ we may analogously obtain a substitution $i \in S(\top\bot\neg)$ such that $iA$ is equivalent to $B$ by replacing $\neg(p_0 \wedge \neg p_0)$ with $\top$ and $p_0 \wedge \neg p_0$ with $\bot$ in the above definition. □

**Proposition 4.5.4** $\mathsf{Grz} \notin L(S(\neg\wedge))$ *and* $\mathsf{Grz} \notin L(S(\top\bot\neg))$.

***Proof*** The proof is completely parallel to the proof of Proposition 4.5.2, except that when we assume that $i$ is a $S(\neg\wedge)$ or $S(\top\bot\neg)$ substitution and that $v(i(p \to \Box p)) = 0$ we may infer that $i$ is neither tautologous nor contradictory and can then apply Lemma 4.5.3 to obtain the relevant substitution $j$. □

We end our discussion of these non-modal substitution classes by relating them to Kripke semantics. For the substitution class $H$ we will work with the class of partial function frames $(W, \subseteq)$ where:

$W := X \rightharpoonup \{\top, \bot\}$ (the set of partial functions from a finite set $X$ to a two element set).

We call the logic of these frames *the logic of finite partial functions*.

Recall that the Kripke frame $(W_H, R_H)$ associated with $H$ is defined by letting $W_H = H$ and $R_H = \{(i, j \circ i) \mid i, j \in H\}$, and the valuation $V$ on $(W_H, R_H)$ associated with a $H$-valuation $v$ is defined by $V(i, p_k) = v(ip_k)$.

**Proposition 4.5.5** *For any finite set $X$ there is a surjective p-morphism, $f$, from the Kripke frame associated with $H$, $(W_H, R_H)$, to the finite partial function frame $(X \rightharpoonup \{\top, \bot\}, \subseteq)$.*

*Moreover, for any $H$-valuation $v$ and letters $p_1 \ldots p_n$ where $n = |X|$, there exists a valuation $U$ for which the truth-values of $p_1 \ldots p_n$ are preserved by $f$:*

$V(i, p_k) = U(f(i), p_k)$ *for* $k = 1 \ldots n$

*for $V$ the valuation associated, defined by $V(i, p_k) = v(ip_k)$.*

***Proof*** For convenience we will let $X = \{p_1 \ldots p_n\}$. Define a function $f : H \to (\{p_1 \ldots p_n\} \rightharpoonup \{\top, \bot\})$ as follows:

$$f(i)(p_k) = \begin{cases} \text{undefined} & \text{if } ip_k = p_k \\ ip_k & \text{otherwise} \end{cases}$$

It remains to show that $f$ satisfies the two conditions for being a $p$-morphism. To establish the first condition we must show that $f(i) \subseteq f(j \circ i)$ for any $i$ and $j$ in $H$. If $f(i)(p_k)$ is defined and $= \top$ then $ip_k = \top$ and so $jip_k = \top$, and thus $f(j \circ i)(p_k) = \top$. Similarly if $f(i)(p_k)$ is defined and $= \bot$ then $f(j \circ i)(p_k) = \bot$, so $f(i) \subseteq f(j \circ i)$.

The second condition amounts to showing that, for any $i \in H$ and $w \in \{p_1 \ldots p_n\} \rightharpoonup \{\top, \bot\}$, if $f(i) \subseteq w$ then there exists a substitution $j \in H$ such that $f(j \circ i) = w$. $j(p_k)$ may be defined as $w(p_k)$ if $w(p_k)$ is defined, and as $p_k$ otherwise.

Given a $H$-valuation $v$, we can now define a valuation $U$ on $(\{p_1 \ldots p_k\} \rightharpoonup \{\top, \bot\})$:

$$U(w, p_k) = \begin{cases} v(p_k) & \text{if } w(p_k) \text{ is undefined} \\ 1 & \text{if } w(p_k) = \top \\ 0 & \text{if } w(p_k) = \bot \end{cases}$$

To establish that $U$ respects the letters $p_1 \ldots p_n$ we must show that $v(ip_k) = V(f(i), p_k)$ for each $k = 1, \ldots, n$. If $ip_k = p_k$, then $f(i)(p_k)$ is undefined and so by the definition of $V$, $V(f(i), p_k) = v(p_k)$. But since $ip_k = p_k$, $V(f(i), p_k) = v(ip_k)$

as required. Otherwise $ip_k = \top$ or $ip_k = \bot$. In the former case $V(f(i), p_k) = 1$ by definition of $V$, and $v(ip_k) = v(\top) = 1$, so they are identical as required. The latter case is proved in the same manner. □

**Theorem 4.5.6** *$L(H)$ is the logic of finite partial functions.*

***Proof*** We shall show that a sentence is consistent with $L(H)$ iff it is satisfiable in some finite partial function frame.

If $A$ is consistent with $L(H)$ then for some (arbitrary) substitution instance $A'$ and some $H$-valuation, $v$, $v(A') = 1$. Let $p_1 \ldots p_n$ be the letters in $A'$. It is immediate from Proposition 4.5.5 that there is a valuation $U$ over the partial function frame $(\{p_1 \ldots p_n\} \rightharpoonup \{\top, \bot\}, \subseteq)$ such that $U(f(\iota), A') = 1$ and so $A'$ is satisfiable in a finite partial function frame. Thus, $A$ must be also consistent in the logical of partial function frames, for otherwise its substitution instances would be inconsistent too.

Now suppose that $A$ is true at some world $w$ in a partial function model $(X \rightharpoonup \{\top, \bot\}, \subseteq, U)$: $U(w, A) = 1$. By Proposition 4.5.5 there exists a surjective $p$-morphism from $(W_H, R_H)$ to $(X \rightharpoonup \{\top, \bot\}, \subseteq)$. Choose some substitution $k \in H$ such that $f(k) = w$.

We now follow the proof of Proposition 4.3.11. Let $v$ be any $H$-valuation and $V$ the associated valuation on $(W_H, R_H)$. We can construct a substitution $i$ (which need not necessarily belong to $H$) such that for any $j \in H$, $V(j, ip_m) = U(f(j), p_m)$. Then, because $f$ is a $p$-morphism, $V(j, iB) = U(f(j), B)$ for any sentence $B$ (see Proposition 4.3.11). So in particular, $V(k, iA) = U(f(k), A) = U(w, A) = 1$. Thus some substitution instance of $A$, namely $kiA$, is true in a $H$-valuation, namely $v$, and thus $A$ is consistent in $L(H)$.

Here is how to find such a substitution $i$. For each $w : X \rightharpoonup \{\top, \bot\}$ define a sentence

$$C_w := \bigwedge_{w(p_m)=\top} \Box p_m \wedge \bigwedge_{w(p_m)=\bot} \Box \neg p_m \wedge \bigwedge_{p_m \notin dom(w)} (\Diamond p_m \wedge \Diamond \neg p_m)$$

with $p_m$ ranging over the letters $p_1 \ldots p_n$. Now consider the substitution:

$$ip_m := \bigvee_{U(w, p_m)=1} C_w$$

Observe that $C_w$ and $C'_w$ are inconsistent in the propositional calculus when $w \neq w'$.

We establish that for any $j \in H$, $V(j, C_{f(j)}) = v(jC_{f(j)}) = 1$. There is a conjunct of $jC_{f(j)}$ for each $p_m$ where $m \in \{1, \ldots, n\}$: we will show that each such conjunct is true. If $f(j)(p_m)$ is undefined, then $jp_m = p_m$ and the relevant conjunct $\Diamond jp_m \wedge \Diamond \neg jp_m$ is true according to $v$. If $f(j)(p_m) = \top$ or $\bot$ then $jp_m = \top$ or $\bot$, respectively, and the relevant conjunct, $\Box jp_m$ or $\Box \neg jp_m$ respectively, is true according to $v$.

Now $V(j, ip_m) = 1$ iff $v(jip_m) = v(j \bigvee_{U(w,p_m)=1} C_w) = 1$ iff for some $w$ such that $U(w, p_m) = 1$, $v(jC_w) = 1$. $v(jC_{f(j)} = 1$ and the $C$s are pairwise incompatible, $w = f(j)$. So the last statement holds iff $U(f(j, p_m) = 1$, as required. □

We turn next to the logic of the non-modal substitution class. Let $C$ denote the set of substitutions of letters for arbitrary non-modal sentences (i.e., $S(\neg\wedge)$). Consider the class of frames $\mathcal{F}_X = (W, \leq)$, where $X$ is a finite set and:

$W := \{(P, a) \mid P \subseteq \{0, 1\}^X, a \in P\}$
$(P, a) \leq (Q, b)$ iff $P \supseteq Q$.

The logic of the substitution class $C$ may similarly be related to the logic of this class of frames.

**Proposition 4.5.7** *For any finite set X there is a surjective p-morphism, f, from the Kripke frame associated with C, $(W_C, R_C)$, to the frame $\mathcal{F}_X$.*

*Moreover, for any C-valuation v and letters $p_1 \ldots p_n$ where $n = |X|$, there exists a valuation U for which the truth-values of $p_1 \ldots p_n$ are preserved by f:*

$V(i, p_k) = U(f(i), p_k)$ *for* $k = 1 \ldots n$

***Proof*** Let $v$ be any $C$-valuation. Without loss of generality we will suppose that $X = \{p_1 \ldots p_n\}$. Given a finite set of letters, $Z$, and $a \in 2^Z$, let $B^a$ be the formula $\bigwedge_{a(p_m)=1} p_m \wedge \bigwedge_{a(p_m)=0} \neg p_m$.

Define a function $f : C \to \{(P, a) \mid P \subseteq 2^X, a \in P\}$ as follows:

$$f(i) = (\{a \in 2^X \mid v(\Diamond i B^a) = 1\}, v \circ i \restriction X)$$

It remains to show that $f$ satisfies the two conditions for being a $p$-morphism. To establish the first condition we must show that $f(i) \leq f(j \circ i)$ for any $i$ and $j$ in $C$. Suppose $a$ belongs to the first component of $f(j \circ i)$, so that $v(\Diamond jiB_a) = 1$. So for some $k \in C$, $v(kjiB^a) = 1$. But $k \circ j \in C$, and so $v(\Diamond iB^a) = 1$, which means that $a$ belongs to the first component of $f(i)$. Since $a$ was arbitrary, we have shown that $f(i) \leq f(j \circ i)$.

Now suppose that $f(i) \leq (Q, e)$, where $f(i) = (P, d)$. We wish to find a substitution $j$ such that $f(j \circ i) = (Q, b)$. Let $q_1 \ldots q_r$ be the letters appearing in $ip_1 \ldots ip_n$, and call the members of $2^{\{q_1 \ldots q_r\}}$ (truth-value) assignments (to $q_1 \ldots q_r$). Given $a \in 2^{\{q_1 \ldots q_r\}}$ and a propositional formula $A$ in the letters $q_1 \ldots q_r$ we write $a(A)$ for $A$'s truth-value under the assignment $a$. Let $a \circ i : \{p_1 \ldots p_n\} \to \{0, 1\}$ be the function $p_m \mapsto a(ip_m)$, and let $Y$ be the set $\{a \in 2^{\{q_1 \ldots q_r\}} \mid a \circ i \in Q\}$. (Note that *every* element of $P$ is of the form $a \circ i$ for some $a \in X$.) For any $a \in 2^{\{q_1 \ldots q_r\}}$ there is a corresponding $\top\bot$ substitution, $k_a$, defined on $\{q_1 \ldots q_r\}$ defined by $k_a(q_m) = \top$ if $a(q_m) = 1$ and $k_a(q_m) = \bot$ otherwise. For any propositional formula $A$ in the letters $q_1 \ldots q_r$ it is clear that $v(k_a A) = a(A)$.

Given a surjection $\sigma : 2^{\{q_1 \ldots q_r\}} \to Y$, we can define a substitution on $q_1 \ldots q_r$ (as in Lemma 4.3.9):

$$j(q_m) = \bigvee_{\sigma(a)(q_m)=1} B^a$$

where $a$ ranges over members of $2^{\{q_1 \ldots q_r\}}$. Notice that $\sigma(a)(p_m) = 1$ iff $B^a$ is a disjunct of $j(q_m)$, iff $a(jq_m) = 1$, iff $v(k_a jq_m) = 1$. So we have $\sigma(a)(q_m) = a(jq_m) = v(k_a jq_m)$, for any $a \in 2^{\{q_1 \ldots q_r\}}$ and $m \in \{1, \ldots, r\}$, and so for any propositional formula $A$ in letters $q_1 \ldots q_r$:

$$\sigma(a)(A) = a(jA) = v(k_a jA)$$

Since $iB^c$ is a propositional formula, it follows that $\sigma(a)(iB^c) = a(jiB^c) = v(k_a jiB^c)$.

We show firstly that $\{a \mid v(\Diamond jiB^a) = 1\} = Q$. If $c \in Q$ then $c = b \circ i$ for some $b \in Y$ and so there exists an $a$ such that $\sigma(a) = b$ since $\sigma$ is surjective. So $v(k_a jiB^c) = \sigma(a)(iB^c) = b(iB^c) = c(B^c) = 1$. So for any $c \in Q$, $v(\Diamond jiB^c) = 1$, and hence $Q \subseteq \{c \in 2^X \mid v(\Diamond jiB^c) = 1\}$. Conversely, if $v(\Diamond jiB^c) = 1$ then for some substitution $k$, $v(kjiB^c) = 1$. Indeed, since $jiB^c$ is a propositional formula, we may assume without loss of generality that $k$ is a $\top\bot$-substitution $k_a$ for some $a \in X$. So $1 = v(k_a jiB^c) = \sigma(a)(iB^c) = (\sigma(a) \circ i)(B^c)$ and, since $\sigma(a) \in Y$, $\sigma(a) \circ i \in Q$. But $(\sigma(a) \circ i)(B^c) = 1$ iff $\sigma(a) \circ i = c$, $c \in Q$.

We show next that would $v(jiB^e) = 1$, so that $f(j \circ i) = (Q, e)$, as required of a $p$-morphism. Since $e \in Q$, there exists a substitution $k$ such that $v(kjiB^e) = 1$. Moreover, since $jiB^e$ is a propositional formula, any substitution $k'$ such that $v(kq_m) = v(k'q_m)$ will also be such that $v(k'jiB^e) = 1$. So we may assume without loss of generality that $k$ is a substitution that maps $q_m$ to itself or $\neg q_m$ for $m = 1 \ldots r$. We claim that $k \circ j$ is the required substitution. By construction, $v(kjiB^e) = 1$, but also, since we have replaced literals for literals, $\{a \mid v(\Diamond kjiB^a) = 1\} = \{a \mid v(\Diamond jiB^a) = 1\} = Q$.

Finally, define a valuation $U$ on $\mathcal{F}_X$ where, for $m = 1 \ldots n$:

$$U((P, d), p_m) = d(p_m)$$

$U$ may be arbitrarily defined on the remaining letters. It is immediate that $f$ preserves the truth-values of $p_1 \ldots p_n$: $f(i) = (P, v \circ i)$ for some $P$, and so $V(i, p_m) = v(ip_m) = v \circ i(p_m) = U((P, v \circ i), p_m)$. □

**Theorem 4.5.8** *$L(C)$ is the logic of the frames $\mathcal{F}_X$ for finite sets $X$.*

***Proof*** We shall show that a sentence is consistent with $L(C)$ iff it is satisfiable in some frame $\mathcal{F}_X$.

If $A$ is consistent with $L(C)$ then for some (arbitrary) substitution instance $A'$ and some $C$-valuation, $v$, $v(A') = 1$. Let $X = \{p_1 \ldots p_n\}$ be the set of letters in $A'$. It is immediate from Proposition 4.5.7 that there is a valuation $U$ over the frame $\mathcal{F}_X$ such that $U(f(\iota), A') = 1$ and so $A'$ is satisfiable in $\mathcal{F}_X$. $A$ must be also consistent in the logical of partial function frames, for otherwise its substitution instances would be inconsistent too.

Now suppose that $A$ is true at some world $(P, d)$ in a model $(\mathcal{F}_X, U)$: $U((P, d), A) = 1$. By Proposition 4.5.7 there exists a surjective $p$-morphism from $(W_C, R_C)$ to $\mathcal{F}_X$. Choose some substitution $k \in C$ such that $f(k) = (P, d)$.

We now follow the proof of Proposition 4.3.11. Let $v$ be any $C$-valuation and $V$ the associated valuation on $(W_C, R_C)$. We can construct a substitution, $i$ (which need not necessarily belong to $C$), such that for any $j \in H$, $V(j, ip_m) = U(f(j), p_m)$. Then, because $f$ is a $p$-morphism, $V(j, iB) = U(f(j), B)$ for any sentence $B$ in letters $p_1 \ldots p_n$. So in particular, $V(k, iA) = U(f(k), A) = U((P, d), A) = 1$. Thus some substitution instance of $A$, namely $kiA$, is true in a $C$-valuation, namely $v$, and thus $A$ is consistent in $L(C)$.

For each world $(P, d)$ of $\mathcal{F}_X$ define a sentence:

$$C_{P,d} := \bigwedge_{a \in P} \Diamond B^a \wedge \bigwedge_{a \in 2^X \setminus P} \neg \Diamond B^a \wedge B^d$$

where, as before, $B^a = \bigwedge_{a(p_m)=1} p_m \wedge \bigwedge_{a(p_m)=0} \neg p_m$ for $m$ ranging from 1 to $n$. Now consider the substitution:

$$ip_m := \bigvee_{U((P,d),p_m)=1} C_{P,d}$$

Observe that $C_{P,d}$ and $C_{P',d'}$ are inconsistent in the propositional calculus when $(P, d) \neq (P', d')$.

I can be shown that for any $j \in C$, $V(j, C_{f(j)}) = v(jC_{f(j)}) = 1$. In fact, this is trivial from the definition of $f$: the conjuncts of the form $\Diamond jB^a$ in $C_{f(j)}$ are defined as those where $v(\Diamond jB^a) = 1$, the conjuncts of the form $\neg \Diamond jB^a$ in $C_{f(j)}$ are defined as those where $v(\Diamond jB^a) \neq 1$, and the final conjunct is $B^{v \circ j}$ and $v \circ j(B^{v \circ j}) = 1$ (since $a(B^a) = 1$ for any $a \in 2^X$).

Now $V(j, ip_m) = 1$ iff $v(jip_m) = v(j \bigvee_{U((P,d),p_m)=1} C_{P,d}) = 1$ iff for some $(P, d)$ such that $U((P, d), p_m) = 1$, $v(jC_{P,d}) = 1$. $v(jC_{f(j)} = 1$ and since the $C$s are pairwise incompatible, $(P, d) = f(j)$. So the last statement holds iff $U(f(j), p_m) = 1$, as required. □

The final non-modal substitution class we will consider is $K$. In Sect. 4.4.2 we showed that the pre-validities of $K$ were identical to the $C$-valid sentences—the sentences valid according to Carnap's interpretation of propositional modal logic. In Thomason (1973) an axiom system extending S5 is presented, and it is shown to be complete with respect to the $C$-valid sentences. It consists of the result of closing all instances of the following axioms under modus ponens and the rule of necessitation (but *not* the rule of uniform substitution)[21]:

[21] This system is closely related to S13 from Cocchiarella (1974), the main difference being that Cocchiarella's system does not have a primitive notion of negation.

PC Any instance of a propositional tautology.
K $\Box(A \to B) \to \Box A \to \Box B$
T $\Box A \to A$
5 $\neg\Box A \to \Box\neg\Box A$
Log $\neg\Box A$ when $A$ is a propositional formula that is not tautologous.

It follows from Proposition 4.4.5 and Thomason's completeness result that this system completely axiomatizes the pre-validities of $K$.

What of the *validities* of the substitution class $K$, $L(K)$? In light of the fact that $C$-validity is not closed under the rule of substitution, Cresswell (2013) has proposed that we treat a sentence as valid according to the Carnapian interpretation of $\Box$ only if all of its substitution instances are $C$-valid. Thus Cresswell's proposed notion of validity stands to Carnap's as the present notion of validity with respect to a substitution class stands to pre-validity (as in Definitions 4.2.7 and 4.2.8). Cresswell proves that the validities of Carnap's theory are exactly the theorems of S5. Thus we have

**Corollary 4.5.9** $L(K) = \mathsf{S5}$.

***Proof*** From Proposition 4.4.5 and Theorem 3.3 of Cresswell (2013). □

### 4.5.2 The Logic of the Full Substitution Class

Section 4.3.1 established the existence of a valuation for the full substitution class $S(\neg \wedge \Box)$, and so we know that the logic of this class is non-trivial. While we do not know what the logic of the full substitution class is, we will show in this section that, given a variant of Friedman's conjecture (Conjecture 4.3.2 above), it is exactly Med, and that even if this conjecture is false, the logic of the full substitution class contains Sub, a distinctive principle belonging to Med.

The uniqueness conjecture tells us that for any $S(\neg \wedge \Box)$-valuation $v$, $v(\Box A) = 1$ iff $A \in \mathsf{Med}$. We may now see how this conjecture settles the logic of the full substitution class:

**Proposition 4.5.10** *Given the uniqueness conjecture,* $L(S(\neg \wedge \Box)) = \mathsf{Med}$.

***Proof*** Suppose $A \in \mathsf{Med}$. So for any $S(\neg \wedge \Box)$-valuation $v$, $v(\Box A) = 1$ by the conjecture and Theorem 4.3.8. So $v(i\,A) = 1$ for every substitution $i \in S(\neg \wedge \Box)$. Since $v$ was arbitrary, $A$ is valid. Conversely, suppose that $A$ is valid, so that $v(i\,A) = 1$ for every substitution $i \in S(\neg \wedge \Box)$ and $S(\neg \wedge \Box)$-valuation $v$. By Theorem 4.3.8 there is at least one such $v$, and by the condition for necessity, $v(\Box A) = 1$. Moreover, according to the valuation constructed in Theorem 4.3.8 $v(\Box A) = 1$ only if $A \in \mathsf{Med}$. □

Note that we only appealed to the conjecture in proving one of the two directions, allowing us to establish the following without assuming the conjecture.

**Proposition 4.5.11** $\mathsf{S4M} \subseteq L(S(\neg \vee \Box)) \subseteq \mathsf{Med}$.

This rules out some extensions of S4M, but obviously leaves open any modal logic between S4M and Med.

One distinctive feature of Med is that it contains something we will call the "subset principle". To state this principle we will begin by describing some formulas $D_Y$ for $Y \in P^0(X)$ that characterize the worlds of certain sorts of Medvedev models of the form $(P^0(X), \supseteq, V)$. Write $Z \subseteq^0 Y$ ($Z \subset^0 Y$) to mean that $Z$ is a non-empty (proper) subset of $Y$. To each set, $X = \{1, \ldots, n\}$, associate in some canonical way a propositional partition $A_1 \ldots A_n$. For non-empty $Y \subseteq X$, we will define a formula $D_Y^X$ (or simply $D_Y$ when $X$ is clear from context). When $|Y| = 1$:

$$D_{\{m\}} = \Diamond\Box A_m \wedge \bigwedge_{k \neq m} \neg\Diamond\Box A_k$$

When $D_Z$ is defined for $|Z| \leq k$ and $|Y| = k + 1$:

$$D_Y = \bigwedge_{Z \subset^0 Y} \Diamond D_Z \wedge \Box(\bigwedge_{Z \subset^0 Y} \Diamond D_Z \vee \bigvee_{Z \subset^0 Y} D_Z).$$

The subset principle is then[22]:

Sub $\quad \bigvee_{Y \in P^0(X)} D_Y$

Observe that Sub implies the formula $\Box(\bigwedge_{Z \subset^0 Y} \Diamond D_Z \vee \bigvee_{Z \subset^0 Y} D_Z)$ for each non-empty $Y \subseteq X$, and indeed, these formulas provide an equivalent formulation of Sub.

**Proposition 4.5.12** *Every instance of* Sub *is in* Med.

***Proof*** Consider a Medvedev model $(P^0(Y), \supseteq, V)$, and for each $w \in P^0(Y)$ let $f(w) = \{k \in X \mid w \Vdash \Diamond\Box A_k\}$ for $X = \{1, 2, \ldots, n\}$.

Claim: $w \Vdash D_{f(w)}$

The proof is by induction on the cardinality of $f(w)$. If $|f(w)| = 1$ then $f(w) = \{m\}$ for some $m \in X$, so $w \Vdash \Diamond\Box A_m$ and $w \Vdash \neg\Diamond\Box A_k$ for $m \neq k$. That is, $w \Vdash D_{\{m\}}$.

Suppose the claim is true when $|f(w)| \leq k$. We must now show:

1. $w \Vdash \Diamond D_Z$ for each $Z \subset f(w)$.
2. $w \Vdash \Box(\bigwedge_{Z \subset^0 f(w)} \Diamond D_Z \vee \bigvee_{Z \subset^0 f(w)} D_Z)$.

For 1, note that for any $Z \subset^0 f(w)$, $w$ sees at least a world $v$ such that $f(v) = Z$, namely $\{a \in w \mid \exists k \in Z, \{a\} \Vdash A_k\}$. By the inductive hypothesis $v \Vdash D_{f(v)}$ where $f(v) = Z$, and so $w \Vdash \Diamond D_Z$.

For 2, suppose that $w \supseteq v$. If $f(v) = f(w)$ then $w \Vdash \bigwedge_{Z \subset^0 f(w)} \Diamond D_Z$ by the previous argument. Otherwise $f(v) \subset f(w)$, and by the inductive hypothesis $v \Vdash D_{f(v)}$, and so $v \Vdash \bigvee_{Z \subset^0 f(w)} D_Z$. □

---

[22] It can be see to be a generalization of an axiom schema from Holliday (2017) and Hamkins et al. (2015), where $1 < k \leq m$:

$$(\bigwedge_{i \leq m} \Diamond\Box A_i \wedge \neg\Diamond \bigvee_{i \neq j} A_i \wedge A_j) \rightarrow \Diamond(\bigwedge_{i \leq k-1} \Diamond\Box A_i \wedge \bigwedge_{k \leq j \leq m} \neg\Diamond\Box A_j)$$

Our axiom is strictly stronger than this axiom even in the presence of S4.

Note that $f$ is a p-morphism from the present Medvedev frame on $P^0(Y)$ to the frame on $P^0(X)$.

It may also be shown that all instances of Sub are valid for the full substitution class:

**Proposition 4.5.13** $\mathsf{Sub} \in L(S(\neg \wedge \Box))$.

***Proof*** Let $X = \{1, \ldots, n\}$, and let $v$ be an arbitrary valuation of the full substitution class. We will show that $v(i\, D_{f(i)}) = 1$. It follows that $v(i \bigvee_{Y \subseteq X} D_Y) = 1$ for every substitution $i$, securing the validity of Sub.

Base: suppose $|f(i)| = 1$. So $f(i) = \{m\}$, which means that $v(\Diamond\Box i\, A_m) = 1$ and $v(\Diamond\Box i\, A_r) = 0$ for $r \neq m$.

Inductive Step: suppose the inductive hypothesis holds when $|f(i)| \leq m$, and suppose that $|f(i)| = m + 1$.

We must show

(i) $v(\Diamond i\, D_Z) = 1$ for each $Z \subset f(i)$ and
(ii) $v(\Box(\bigwedge_{Z \subset f(i)} \Diamond i\, D_Z \vee \bigvee_{Z \subset f(i)} i\, D_Z)) = 1$.

For (i), we may appeal to Lemma 4.3.9 to obtain a $j$ such that $f(j \circ i) = Z$. Since $|f(j \circ i)| \leq n$ we may apply the inductive hypothesis and conclude that $v(ji\, D_{f(j \circ i)}) = 1$, and hence that $v(\Diamond i\, D_{f(j \circ i)}) = 1 = v(\Diamond i\, D_Z)$.

We now show (ii). For any $j$, $f(j \circ i) \subseteq f(i)$. If $f(j \circ i) = f(i)$ then $v(\Diamond ji\, D_Z) = 1$ for each $Z \subset f(i) = f(j \circ i)$ by repeating the reasoning for part (i) (using $j \circ i$ instead of $i$). If $f(j \circ i) \subset f(i)$ then $v(ji\, D_{f(j \circ i)}) = 1$ by the inductive hypothesis. So for any substitution $j$, $v(j(\bigwedge_{Z \subset f(i)} \Diamond i\, D_Z \vee \bigvee_{Z \subset f(i)} i\, D_Z)) = 1$ and so $v(\Box(\bigwedge_{Z \subset f(i)} \Diamond i\, D_Z \vee \bigvee_{Z \subset f(i)} i\, D_Z)) = 1$. □

Thus, we have a tight bound on the logic of the full substitution class:

$$\mathsf{S4MSub} \subseteq L(S(\neg \wedge \Box)) \subseteq \mathsf{Med}.$$

There are, however, further valid principles of Med that we have not been able to verify to be valid for the full substitution class. One is:

(A1) $\quad D_Y \rightarrow \Diamond(D_Y \wedge \bigwedge_{Z \subseteq^0 Y} \bigvee_k \Box(D_Z \rightarrow A_k))$

Where $A_1 \ldots A_r$ is an arbitrary partition of propositional formulas.

(A1) may be seen to be valid as follows. Let $(P^0(Y), \supseteq, V)$ be a model over a Medvedev frame. Suppose $D_Z$ is true at $w$, where $Z \subseteq X = \{1, \ldots, n\}$. The formulas $A_1 \ldots A_n$ partition the terminal worlds $\{\{m\} \mid m \in w\}$ according to which $A_k$ they make true. For each formula $A_k$ of $A_1 \ldots A_n$ such that $w \Vdash \Diamond\Box A_k$, pick a single representative $m_k \in w$, such that $\{m_k\} \Vdash A_k$. It is clear that $w$ sees $\{m_1 \ldots m_n\}$, and that for any $Z \subseteq \{1, \ldots, n\}$, $D_Z$ is true only at the corresponding subset of $\{m_1, \ldots, m_n\}$: $\{m_k \mid k \in Z\}$. Because each $D_Z$ is true in exactly one world seen by $\{m_1, \ldots, m_n\}$, it follows that $\{m_1, \ldots, m_n\} \Vdash D_Y \wedge \Box \bigwedge_{Z \subseteq^0 Y} \bigvee_k \Box(D_Z \rightarrow B_k)$.

Another is a rule under which Med is closed:

(R1) $\quad$ If $\vdash (D^X_X \wedge \bigwedge_{Z \subseteq^0 X} \bigvee_k \Box(D^X_Z \rightarrow S_k)) \rightarrow C$ for for every finite set $X$, then $\vdash C$

where $S_k$ ranges over all possible consistent conjunctions of literals in the letters appearing in $C$ and $D_X$. Indeed, it may be shown, without too much difficulty, that Med is the smallest modal logic containing S4, Sub and closed under (R1). Thus if we could show that $\Delta_v$ is closed under (R1) whenever $v$ is a valuation of the full substitution class, the uniqueness conjecture would be solved.

We may also use the fact that Sub belongs to the logic of the full substitution class to fully resolve Humberstone's second question: the logic of the full substitution class is different from the logic of $H$.

**Proposition 4.5.14** *$L(H)$ does not contain all instances of Sub.*

***Proof*** Let $A_1 \ldots A_4$ be the four way partition of propositional formulas: $p_1 \wedge p_2$, $p_1 \wedge \neg p_2$, $\neg p_1 \wedge p_2$, and $\neg p_1 \wedge \neg p_2$.

In any $H$-valuation, it is easy to see that $p_1 \wedge p_2$, $p_1 \wedge \neg p_2$, $\neg p_1 \wedge p_2$, and $\neg p_1 \wedge \neg p_2$ are all possibly necessary by considering the four possible ways of substituting $\top$ and $\bot$ for $p_1$ and $p_2$. Thus $v(D_Y) = 0$ for any proper subset $Y$ of $\{1, 2, 3, 4\}$ and valuation $v$. So if Sub were valid, then $v(D_{\{1,2,3,4\}}) = 1$; and so, in particular, $v(\Diamond D_{\{1,4\}}) = 1$ by the definition of $D_{\{1,2,3,4\}}$.

But $\Diamond D_{\{1,4\}}$ implies $\Diamond\Box(p_1 \wedge p_2)$ and $\Diamond\Box(\neg p_1 \wedge \neg p_2)$, and also $\neg\Diamond\Box(p_1 \wedge \neg p_2)$ and $\neg\Diamond\Box(\neg p_1 \wedge p_2)$. Yet any $H$ substitution according to which $p_1 \wedge p_2$ is possibly necessary and $\neg p_1 \wedge \neg p_2$ is possibly necessary, $p_1$ and $p_2$ must be mapped to themselves, and not to $\top$ or $\bot$, since it must remain possible to change the truth values of both $p_1$ and $p_2$. But relative to any substitution in which $p_1$ and $p_2$ are mapped to themselves, $\neg p_1 \wedge p_2$ and $p_1 \wedge \neg p_2$ must also be possibly necessary, since $p_1$ and $p_2$ may be replaced with $\top$ or $\bot$ at will. □

We end this section with a weaker conjecture. It is interesting to note that in the $S(\neg \wedge \Box)$-valuations that we constructed by setting $v(\Box A) = 1$ iff $A \in$ Med, every possibility is witnessed by a substitution of sentence letters with sentences of modal degree 2 (the sentences $D^i_p$)—for short, a substitution of modal degree 2. So if $v$ is such a valuation, and if $v(\Diamond A) = 1$ then there is some substitution $i$ of modal degree 2 such that $v(iA) = 1$. (Note that the substitutions of modal degree 2 do not themselves form a substitution class, since they are not closed under composition, despite the fact that they are well-behaved in this context.) This suggests the following more general claim:

**Conjecture 4.5.15** If $v$ is a $S(\neg \wedge \Box)$-valuation and $v(\Diamond A) = 1$, then $v(iA) = 1$ for some substitution of modal degree 2.

Given the above observation, this conjecture clearly follows from Conjecture 4.3.2, but it might more easily be tackled directly.

## 4.6 Other Theories of Logical Necessity

In the introduction we considered a couple of alternatives to McKinsey's constraint. The first replaces the substitutional analysis of logical truth with an arbitrary logic $\Delta$. The propositional analogue of logical truth, so conceived, is an interpretation of the modal operator $\Box$ under which a sentence $\Box A$ is true just in case $A$ is a member of $\Delta$. In the second, we replaced the Bolzanoean substitutional analysis with a Tarski-style analysis in a higher-order logic, in which a sentence $A(c_1 \ldots c_n)$ is logically true just in case the $\forall x_1 \ldots x_n . A(x_1 \ldots x_n)$ is true, where $c_1 \ldots c_n$ enumerate the non-logical constants appearing in $A$.[23] The corresponding constraint on logical necessity can then by stated by a single biconditional:

$$\Box A(c_1 \ldots c_n) \leftrightarrow \forall x_1 \ldots x_n . A(x_1 \ldots x_n)$$

In this section we will investigate both of these options.

### *4.6.1 The Tarskian Constraint*

According to the Tarskian analysis, one can express the logical truth of a sentence of a propositional modal logic, such as $\Box p \rightarrow p$, with a single sentence of propositionally quantified modal logic, in this case $\forall X(\Box X \rightarrow X)$.[24] More generally, the logical truth of a closed sentence $A(p_1 \ldots p_n)$ in sentential constants $p_1 \ldots p_n$ should be equivalent to the truth of the quantified sentence $\forall X_1 \ldots X_n . A(X_1 \ldots X_n)$; and so one might articulate the idea that a propositional operator expresses the worldly analogue of logical truth under this conception by means of the following schema:

$$\Box A(p_1 \ldots p_n) \leftrightarrow \forall X_1 \ldots X_n . A(X_1 \ldots X_n)$$

This is clearly a restriction of what we earlier called the Tarskian Constraint to a sublanguage of higher-order logic.

We will find it helpful to consider the dualized form of this schema:

$$\Diamond A(p_1 \ldots p_n) \leftrightarrow \exists X_1 \ldots X_n . A(X_1 \ldots X_n)$$

employing standard definitions for $\exists$ and $\Diamond$.

---

[23] Tarski's theory of logical truth was originally formulated in a higher-order logic, as opposed to the set-theoretic model theory it has come to be associated with. The definition used here can be found earlier in Bernays and Schönfinkel (1928, p. 347), and a related definition is found in Hilbert and Ackermann (1928, Chap. IV, Sect. 1). See also the notion of 'Metaphysical Universality' from Williamson (2013).

[24] See (Fritz, 2024).

This schema may be regarded as expressing a principle of recombinatorialism with respect to the atomic propositions, a position once popular among the early logical atomists. For let us suppose that the sentential constants $p_1, \ldots, p_n$ of the language express atomic propositions. Since $A(X_1 \ldots X_n)$ contains propositional variables, not sentential constants, it expresses a relation among propositions defined in purely logical terms. So the right-to-left direction of the equivalence says that the particular atomic propositions $p_1, \ldots, p_n$ can instantiate any logical pattern that is in fact instantiated by some propositions, and the left-to-right direction tells us that these exhaust the logical patterns that the atomic propositions can instantiate. For instance, since there are truths and falsehoods, two instances of the schema ($\Diamond p_0 \leftrightarrow \exists X.X$ and $\Diamond \neg p_0 \leftrightarrow \exists X.\neg X$) tell us that every atomic proposition must be contingent. Similarly, since there are logically necessary propositions and logically false propositions, we may infer that every atomic proposition is possibly necessary and possibly necessarily false: $\Diamond \Box p_0$ and $\Diamond \Box \neg p_0$.[25] It is also worth noting that for this very reason we should not expect the schema itself to be necessary: we have already established that it is possible that $p_0$ is necessary, and also that certain instances of the schema imply that $p_0$ is contingent and so if all instances of the schema were necessary it would be possible for $p_0$ to be both necessary and contingent.

Here we show that there are indeed interpretations of propositionally quantified modal logic under which the schema is true. In the rest of this section, we let $\mathcal{L}$ refer to the language of propositionally quantified modal logic, with the logical constants $\neg, \wedge, \forall, \Box$ and with infinitely many sentence letters. This language augments the language of propositional modal logic with an infinite set of variables that may occupy sentence position, $X_1, X_2, \ldots$, and a quantifier $\forall$ that can bind sentence variables. In addition to the usual syntactic clauses for propositional modal logic we also stipulate that sentence variables are formulas, and that $\forall X_k.A$ is a formula whenever $A$ is a formula.

We consider the "full" Kripke models for this language in which the propositional quantifiers range over all sets of worlds (see Fine, 1970; Fritz, 2024 def. 2.3.1). Given an ordinary Kripke model for propositional modal logic, $(W, R, V)$, we interpret an arbitrary formula of propositionally quantified modal logic as follows. Let $g$ and $g'$ range over variable assignments mapping each propositional variable $X_k$ to a subset of $W$:

- $V^g(w, p_k) = V(w, p_k)$
- $V^g(w, X_k) = 1$ iff $w \in g(X_k)$
- $V^g(w, A \wedge B) = min(V^g(w, A), V^g(w, B))$
- $V^g(w, \neg A) = 1 - V^g(w, A)$

[25] The existence of truths and falsehoods may be derived given the usual axioms for the propositional quantifiers. E.g., existential generalization lets one move from $\top$ to $\exists X.X$ and $\neg\bot$ to $\exists X.\neg X$. The existence of logically necessary propositions and logically false propositions can also be derived directly: $\Box\forall X(X \vee \neg X) \leftrightarrow \forall X(X \vee \neg X)$ is an instance of the schema (where $n = 0$), and since the usual quantificational axioms secure $\forall X(X \vee \neg X)$, we may infer $\Box\forall X(X \vee \neg X)$ from which we may conclude $\exists X\Box X$ by existential generalization. The existence of logically false propositions follows by a similar argument.

- $V^g(w, \Box A) = 1$ iff $V^g(w', A) = 1$ for every $w'$ such that $Rww'$.
- $V^g(w, \forall X_k A) = 1$ iff $V^{g'}(w, A) = 1$ for every $g'$ for which $g$ and $g'$ agree on every variable except possibly on $X_k$.

We may now construct a model of the Tarskian constraint.[26]

**Theorem 4.6.1** *There is a model of the Tarskian constraint.*

***Proof*** Consider the Kripke frame:

- $W := \mathbb{N}^{<\omega}$ (finite sequences of naturals)
- $Rww'$ iff $w$ is a (proper or improper) initial segment of $w'$.

The frame can be visualised as an infinitary tree, that has the empty sequence as its root, and that branches infinitely many times at each node. Enumerate the sentences satisfiable at the root of this frame – $A_1, A_2, A_3 \ldots$. For each $n$, let $(W, R, V_n)$ be a Kripke model over this frame in which $A_n$ is true at the root. We paste these models together to make a single model over $(W, R)$ as follows:

- $V((), p_i)$ may be set arbitrarily
- $V((n_1, \ldots, n_k), p_i) = V_{n_1}((n_2, \ldots, n_k), p_i)$ when $k \geq 1$

It is clear by construction that $A_n$ is true at the world $(n)$, and thus $\Diamond A_n$ is true at the root world $()$ for every $n$. Let $\mathcal{F} \uparrow w$ be the subframe of $\mathcal{F}$ generated by $w$. Then it is also clear that $(W, R) \uparrow w$ is isomorphic to $(W, R)$ for any world $w \in W$.

Now if $\forall X_1 \ldots X_n . A(X_1 \ldots X_n)$ is true at the root world $()$ then $A$ is valid over the frame $(W, R)$. But since for any world $w \in W$, the generated frame $(W, R) \uparrow w$ is isomorphic to $(W, R)$, it follows that $A$ is valid over the frame $(W, R) \uparrow w$, and so true at the submodel of $(W, R, V)$ generated by $w$. This means that $A$ is true at $w$ in $(W, R, V)$, and since $w$ was arbitrary, $\Box A$ is true at the base world $()$. Conversely, if $\Box A$ is true at $()$, then $A$ must be true at $(n)$ for each $n \in \mathbb{N}$, and thus $\neg A$ is not satisfiable in $(W, R)$. So $\forall X_1 \ldots X_n . A[X_1/p_1 \ldots X_n/p_n])$ is true at $()$. □

We end with a conjecture that we have not been able to resolve:

**Conjecture 4.6.2** There exists a model of the Tarskian constraint in $\mathcal{L}$ in which the propositional quantifiers are interpreted substitutionally.

We may make the conjecture precise as follows: there exists a valuation $v : \mathcal{L} \rightarrow \{0, 1\}$ such that (i) $v(\neg A) = 1 - v(A)$, (ii) $v(A \wedge B) = min(v(A), v(B))$, (iii) $v(\forall p . A) = 1$ iff $v(A[B/p]) = 1$ for every closed sentence $B$ of $\mathcal{L}$ and (iv) $v(\Box A(p_1 \ldots p_n) \leftrightarrow \forall X_1 \ldots X_n . A(X_1 \ldots X_n)) = 1$ for every sentence $A$ in the letters $p_1, \ldots, p_n$.

Given the substitutional interpretation of the quantifiers, the McKinseyan and Tarskian requirements coincide: it is easily verified that the conjecture is true if and

[26] This sort of construction can be generalized to all of higher-order logic—for details see the appendix of Bacon (2020). As opposed to here, the construction there is presented in terms of 'metaphysical substitutions' (see Bacon, 2019 and also Fine, 1977) as opposed to Kripke models.

only if there exists a valuation $v : \mathcal{L} \to \{0, 1\}$ satisfying (i)–(iii) that additionally satisfies McKinsey's requirement for the full substitution class of this language: $v(\Box A) = 1$ iff $v(i\,A) = 1$ for every substitution $i$ mapping sentence letters to arbitrary closed sentences of $\mathcal{L}$. The substitutional interpretation of the quantifiers corresponds in a natural way to the idea found in early Wittgenstein and Russell that every proposition may ultimately be analyzed in terms of the logically atomic propositions and logical operations.

### 4.6.2 The Metalogical Constraint

We observed previously that in the valuation constructed in Theorem 4.3.8, $v(\Box A) = 1$ if and only if $A$ was a member of a certain modal logic, Med. Let $\Delta_v = \{A \mid v(\Box A) = 1\}$. Then we have the following general result:

**Proposition 4.6.3** *If $v$ is a $S(\neg \wedge \Box)$-valuation then $\Delta_v$ is a modal logic extending S4M.*

***Proof*** Propositions 4.4.1 and 4.4.3 already establish that whenever $A$ is an instance of K, T, 4 or M then $A$ is valid in $v$, i.e., $v(i\,A) = 1$ for every substitution $i \in S(\neg \wedge \Box)$ and, since $v$ is a $S(\neg \wedge \Box)$-valuation, it follows that $v(\Box A) = 1$, and so $A \in \Delta_v$. Since every instance of 4 is true in $v$, $\Delta_v$ is closed under the rule of necessitation. Finally, $\Delta_v$ is closed under the rule of substitution since if $v(\Box A) = 1$, then $v(\Box\Box A) = 1$, and so $v(\Box i\,A) = 1$ for any substitution $i \in S(\neg \wedge \Box)$. □

Note that for an arbitrary substitution class, $S$, and $S$-valuation $v$, $\Delta_v$ will contain all instances of K, T, 4 and will be closed under the rule of necessitation. However, it may not be a modal logic because it may not be closed under the rule of substitution, but will only be guaranteed to be closed under substitutions in $S$.

It follows that the interpretation of $\Box$ in a $S(\neg \wedge \Box)$-valuation will satisfy an instance of the Meta-Logical Constraint considered in the introduction, according to which the interpretation of logical necessity is identified with a fixed modal logic $\Delta$. We make this precise as follows:

**Definition 4.6.4** (*Meta-valuations*) Let $\Delta$ be a normal modal logic. A function $v : \mathcal{L}(\neg \wedge \Box) \to \{0, 1\}$ is a $\Delta$-valuation if and only if

- $v(A \wedge B) = \min(v(A), v(B))$
- $v(\neg A) = 1 - v(A)$
- $v(\Box A) = 1$ iff $A \in \Delta$

As with the substitutional constraint, it is easily checked that whenever $\Delta$ is a consistent normal modal logic, then 4 must be true in a $\Delta$-valuation, and moreover, the left-to-right direction of McKinsey's constraint is satisfied.

Any modal logic may, of course, be plugged into this definition. But if $\Delta$ is supposed to represent the logical truths and $v$ a possible interpretation of $\Box$ as logical

truth, then we would like the truths under the interpretation to include the logical truths: $v(A) = 1$ whenever $A \in \Delta$. Following Meyer (1971) we say:

**Definition 4.6.5** A modal logic $\Delta$ is coherent iff for every $A \in \Delta$ and $\Delta$-valuation $v$, $v(A) = 1$.

Intuitively, a modal logic is coherent when it accommodates an interpretation of $\Box$ with its own logic. Not every modal logic is coherent. Indeed, we can see that coherent modal logics cannot contain the B axiom (much as McKinsey's constraint rules out the Brouwerian axiom, except in degenerate cases). For suppose, for reductio, that $\Delta$ is a consistent modal logic and that $v$ is a $\Delta$-valuation in which every instance of B holds. For a sentence letter, $p$, $v(p) = 1$ or $v(\neg p) = 1$, so that either $v(\Box\Diamond p) = 1$ or $v(\Box\Diamond\neg p) = 1$. If the former, then $\Diamond p \in \Delta$, and since $\Delta$ is closed under the rule of substitution, $\Diamond\bot \in \Delta$, contradicting the assumption that $\Delta$ is a consistent modal logic. In the latter, we may similarly conclude $\Diamond\neg\top \in \Delta$.

Coherence is equivalent to the disjunction property. Here we will show that it is equivalent to an extended version of the disjunction property, which is satisfied by a normal modal logic $\Delta$ when the following holds for any sentences $A_1 \ldots A_n$ and any non-modal sentence $A_0$:

If $A_0 \vee \Box A_1 \vee \cdots \vee \Box A_n \in \Delta$ then $A_k \in \Delta$ for some $k$ with $0 \leq k \leq n$.

Emil Jerabek (2023) has pointed out to us that the extended disjunction property is indeed equivalent to the disjunction property.

**Proposition 4.6.6** *A normal modal logic is coherent iff it has the extended disjunction property (and thus, iff it has the disjunction property).*

***Proof*** Let $\Delta$ be a normal modal logic. Suppose that $\Delta$ is coherent, and $A_0 \vee \Box A_1 \vee \cdots \vee \Box A_n \in \Delta$. Suppose that for each $k$ with $1 \leq k \leq n$, $A_k \notin \Delta$. It suffices to show that $A_0 \in \Delta$. From our first supposition it follows that $v(A_0 \vee \Box A_1 \vee \cdots \vee \Box A_n) = 1$, for any given $\Delta$-valuation $v$. From our second supposition it follows that $v(\Box A_1) = \cdots = v(\Box A_n) = 0$; and so $v(A_0) = 1$. Since $v$ was arbitrary, $A_0$ must be true in every $\Delta$-valuation, and since $A_0$ is a propositional formula it must be a tautology.

Now suppose that $\Delta$ has the extended disjunction property. We will show that if a formula $A$ is true in some $\Delta$-valuation, $v$, then $A$ is consistent with $\Delta$, from which it follows that $\Delta$ is coherent. Suppose $A$ is a truth functional combination of a formulas of the form $\Box B$ and propositional letters $p_1 \ldots p_n$. Let $\Box B_1 \ldots \Box B_m$ be the true formulas of the form $\Box B$, and $\Box C_1 \ldots \Box C_k$ the false formulas of this form. Finally, let $D$ be the conjunction $\bigwedge_{v(p_i)=1} p_i \wedge \bigwedge_{v(p_i)=0} \neg p_i$. Clearly $v(D) = 1$ so $D$ is truth-functionally consistent. So $\neg D \notin \Delta$. By the definition of a $\Delta$-valuation $B_1 \in \Delta, \ldots, B_m \in \Delta$, and $\neg C_1 \notin \Delta, \ldots, \neg C_k \notin \Delta$. Since $\Delta$ has the extended disjunction property $(\neg D \vee \Box\neg C_1 \vee \cdots \vee \Box\neg C_n) \notin \Delta$. So the set $\{D, \neg\Box\neg C_1, \ldots, \neg\Box\neg C_n, \Box B_1, \ldots, \Box B_m\}$ is consistent with $\Delta$, and since this set entails $A$, $A$ too is consistent with $\Delta$. □

**Remark 4.6.7** (Fine, n.d.) MS has considered a wider class of metavaluations obtained by relaxing the requirement that $\Delta$ be closed under the rule of substitution and/or necessitation. With the rule of substitution relaxed, we may find non-trivial valuations that validate B. (For instance by letting $\Delta$ be the set of $C$-validities; see Sect. 4.4.2.)

If $\Delta$ is not closed under rule of necessitation, the S4 axiom is no longer guaranteed to hold. Urquhart (2010) investigates the result of letting $\Delta$ be the set of substitutions instances of tautologies, yielding an interpretation of logical necessity corresponding to tautologousness, which does not satisfy S4.

Which normal modal logics are coherent? We may see immediately that if $v$ is a valuation of the full substitution class, then $\Delta_v$ is a coherent modal logic.

**Proposition 4.6.8** *If $v$ is a $S(\neg \wedge \Box)$-valuation, $\Delta_v$ is a coherent modal logic.*

***Proof*** Let $A \in \Delta_v$ and let $u$ be any $\Delta_v$ valuation. Since $A \in \Delta_v$, $v(\Box A) = 1$, and so $v(iA) = 1$ for every substitution $i$. We may construct a substitution $j$ such that $v(jA) = u(A)$ by letting $jp_k$ be $p_k$ if $v(p_k) = u(p_k)$ and be $\neg p_k$ otherwise (this is proved by induction). So $u(A) = v(jA) = 1$, as required. □

But many other modal logics are coherent. Here is a fairly general sufficient condition for coherence. Recall that a pointed frame is a Kripke frame $(W, R)$ equipped with a designated world $w_0 \in W$ which bears the ancestral of $R$ to every element of $W$. The logic of a class of pointed frames is the set of sentences true at the designated world in any model over a pointed frame. If $\Delta$ is the logic of a class of Kripke frames $C$, in the ordinary sense, it is also the logic of a class of pointed frames, $\{\mathcal{F} \uparrow w \mid \mathcal{F} \in C, w \text{ a world in } \mathcal{F}\}$, writing $\mathcal{F} \uparrow w$ for the subframe of $\mathcal{F}$ generated by $w$, and $w \uparrow$ to be the set $\{v \mid Rwv\}$ when $R$ is the accessibility relation of $\mathcal{F}$.

**Definition 4.6.9** (*Disjoint p-morphic copies*) A class $C$ of pointed frames is closed under disjoint $p$-morphic copies iff for any pointed frames $\mathcal{F}_1, \ldots, \mathcal{F}_n \in C$, there is another pointed frame $\mathcal{F} \in C$, worlds $w_1 \ldots w_n$ accessible to the designated world of $\mathcal{F}$, and $p$-morphisms $f_i : \mathcal{F} \uparrow w_i \to \mathcal{F}_i$ for $i = 1 \ldots n$ such that $w_i \uparrow \cap w_j \uparrow = \emptyset$ when $i \neq j$, and $f_i(w_i)$ is the designated world of $\mathcal{F}_i$.

**Example 4.6.10** (*Coalesced sums*) Given disjoint pointed Kripke frames $\mathcal{F}_1 \ldots \mathcal{F}_n$, their coalesced sum is the pointed Kripke frame $\mathcal{F} = (W, R, w_0)$ where:

- $W := W_1 \cup \cdots \cup W_n \cup \{w_0\}$
- $R := R_1 \cup \cdots \cup R_n \cup \{w_0\} \times W$

The designated worlds $w_i \in W_i$, and the identity mappings from $\mathcal{F} \uparrow w_i$ to $\mathcal{F}_i$ comprise the relevant $p$-morphisms.

Note that many properties of the component frames are inherited by the coalesced sum. For instance, if $\mathcal{F}_1 \ldots \mathcal{F}_n$ are some combination of the properties of being reflexive, serial or transitive, then $\mathcal{F}$ is also that combination of reflexive, serial or transitive. So these classes are all closed under taking disjoint $p$-morphic copies.

**Example 4.6.11** The singleton class consisting of the frame $(\mathbb{N}^{<\omega}, \leq)$ from Theorem 4.6.1—finite sequences of naturals with the initial subsequence ordering—is closed under disjoint $p$-morphic copies, since it is a countably infinite coalesced sum of isomorphic copies of itself.

**Example 4.6.12** (*The class of Medvedev frames*) The class of Medvedev frames is closed under disjoint $p$-morphic copies.

Suppose $(P^0(X_i), \supseteq)$ for $i = 1...n$ are Medvedev frames. We can suppose, without loss of generality that $X_i \cap Y_j = \emptyset$ when $i \neq j$. Then $(P^0(X_1 \cup \cdots \cup X_n), \supseteq)$ is a Medvedev frame, and the worlds $X_1, \ldots, X_n \in P(X_1 \cup \cdots \cup X_n)$ with the evident identity mappings provide $p$-morphisms into each component $(P^0(X_i), \supseteq)$.

**Example 4.6.13** (*The class of finite partial function frames*) The class of finite partial function frames is closed under disjoint $p$-morphic copies.

Suppose $(X_i \rightharpoonup \{\top, \bot\}, \subseteq)$ are finite partial function frames for $i = 1...n$. We can suppose, without loss of generality that $X_i \cap Y_j = \emptyset$ when $i \neq j$, and that no $X_i$ contains any numbers. Then $((X_1 \cup \cdots \cup X_n \cup \{1, .., n\}) \rightharpoonup \{\top, \bot\}, \subseteq)$ is a finite partial function frame. For each $i = 1...n$, let $w_i$ be any function that is defined everywhere except $X_i$, maps $i$ to $\top$ and maps $j$ to $\bot$ for $j \in 1, \ldots, n$ and $j \neq i$. Due to the last condition, no partial function can be above both $w_i$ and $w_j$ when $i \neq j$. Moreover, the mapping $w \mapsto w \restriction X_i$ is a $p$-morphism from the frame generated by $w_i$ to $((X_1 \cup \cdots \cup X_n \cup \{1, .., n\}) \rightharpoonup \{\top, \bot\}, \subseteq)$.

**Proposition 4.6.14** *If $\Delta$ is characterized by a class of pointed frames that are closed under disjoint $p$-morphic copies, then $\Delta$ is coherent.*

***Proof*** Suppose $\Delta$ the logic of a class of pointed frames $C$ that is closed under disjoint $p$-morphic copies, and suppose that there is some $\Delta$ valuation $u$ such that $u(A) = 1$. We will show that $A$ is true in some model over a frame in $C$. Suppose that $\Box B_1, \ldots, \Box B_n$ are the subformulas of $A$ of the form $\Box B$ such that $v(\Box B) = 0$. So $\neg B_1 \ldots \neg B_n$ are $\Delta$-consistent and thus true in models $M_1 \ldots M_n$ over frames $\mathcal{F}_1 \ldots \mathcal{F}_n \in C$. We know there is a frame $\mathcal{F}$ containing worlds $w_1 \ldots w_n$ and $p$-morphisms $f_i : \mathcal{F} \uparrow w_i \to \mathcal{F}_i$. Let $V$ be any valuation on $\mathcal{F}$ such that whenever $w$ is above $w_i$, $V(w, p_k) = V_i(f_i(w), p_k)$, and $V(w, p_k) = u(p_k)$ when $w$ is the designated world of $\mathcal{F}$. These assignments cannot conflict, since $w_i \uparrow \cap w_j \uparrow = \emptyset$ when $i \neq j$. $A$ is a truth-functional combination of sentence letters and formulas of the form $\Box B$. Now $V(w, p_k) = u(p_k)$ for each sentence letter. If $u(\Box B) = 1$ then $B \in \Delta$ and so $V(w, \Box B) = 1$, since $\mathcal{F} \in C$. If $u(\Box B) = 0$, and $\Box B$ is subformula of $A$, then by construction, $w$ sees a world in which $B$ is false, so $V(w, \Box B) = 0$. Thus, $V(w, A) = u(A)$, as required. □

Appealing to Examples 4.6.10, 4.6.12 and 4.6.13 we may immediately infer:

**Corollary 4.6.15** *K, KT, S4, KD, K4D are coherent.*

**Corollary 4.6.16** *Medvedev logic and the logic of finite partial functions are coherent.*

It also follows that the logic of the infinite tree $(\mathbb{N}^{<\omega}, \leq)$ is coherent, but since this logic is just S4 this observation is already implied by Corollary 4.6.15.

Say that a coherent logic, $\Delta$, is *maximal* iff whenever $\Delta \subseteq \Delta'$ and $\Delta'$ is coherent, $\Delta = \Delta'$. A straightforward application of Zorn's lemma establishes that there exist maximal coherent logics. Here we show that Med is a maximal coherent logic using the formulas described in Sect. 4.3.1.[27]

**Theorem 4.6.17** Med *is a maximal coherent logic.*

***Proof*** Suppose that $\Delta$ is a coherent logic and $\mathsf{Med} \subseteq \Delta$. To show the logics are identical, it suffices to show that anything consistent in Med is consistent in $\Delta$.

Suppose that $C(p_1 \ldots p_m)$, in letters $p_1 \ldots p_m$, is consistent in Med and true at the root $X$ of a Medvedev model $(P^0(X), \subseteq, V)$. Without loss of generality, we may suppose that $X = \{1, \ldots, n\}$. Fix a propositional partition $A_1 \ldots A_n$ and define the formulas $D_Y$, for $Y \subseteq^0 X$ as in Sect. 4.5.2.

Since $A_k$ is a consistent propositional formula, it has a tautologous substitution instance, $A'_k$. Thus $\Diamond\Box A'_k \in \mathsf{KD} \subseteq \Delta$. Since $\Diamond\Box A_k$ has a substitution instance that is consistent in $\Delta$ it is itself consistent in $\Delta$. Since $\Delta$ has the disjunction property, it follows that $\Diamond\Box A_1 \wedge \cdots \wedge \Diamond\Box A_n$ must be consistent in $\Delta$, for otherwise its negation, and thus the disjunction $\Box\Diamond\neg A_1 \vee \cdots \vee \Box\Diamond\neg A_n$ would be in $\Delta$ while none of its disjuncts are.

The formula $\Diamond\Box A_1 \wedge \cdots \wedge \Diamond\Box A_n \rightarrow D_X$ is a consequence of Sub and so belongs to Med and hence also to $\Delta$. Since the antecedent is consistent in $\Delta$, $D_X$ must be consistent as well. So $D_X$ is true at the root of a transitive reflexive Kripke model $(W, R, U)$ which verifies all the theorems of $\Delta$.

Since $\mathsf{Sub} \subseteq \mathsf{Med} \subseteq \Delta$ it follows that for every $w \in W$, there is some $Y \subseteq^0 X$ such that $U(w, D_Y) = 1$, and this $Y$, moreover, is unique because the $D_Y$ are pairwise incompatible. In fact, it is easily seen that $Y$ is $\{k \mid U(w, \Diamond\Box A_k) = 1\}$. For define a function $f : W \rightarrow P^0(X)$ by setting $f(w) = \{k \mid U(w, \Diamond\Box A_k) = 1\}$. This function is a $p$-morphism from the frame $(W, R)$ to $(P^0(X), \subseteq)$. If $Rxy$ then everything possible at $y$ is possible at $x$ since $R$ is transitive, so $f(x) \subseteq f(y)$. If $f(x) \subseteq Y$, then, since $U(x, D_{f(x)}) = 1$, $D_{f(x)}$ has $\Diamond D_Y$ as a conjunct, and so $x$ sees a world $y$ such that $U(y, D_Y) = 1$, and so $f(y) = Y$.

For each letter $p_k$ we may define a formula $D_{p_k}$ as $\bigvee_{V(Y,p_k)=1} D_Y$. It is clear by definition that $U(w, D_{p_k}) = V(f(w), p_k)$, and since $f$ is a $p$-morphism, $U(w, C(D_{p_1} \ldots D_{p_m})) = V(f(w), C(p_1 \ldots p_m))$. We know that $D_X$ is true at the root $w_0$ of $W$, so $f(w_0) = X$ and so $U(w_0, C(D_{p_1} \ldots D_{p_m})) = V(X, C(p_1 \ldots p_m)) = 1$. Thus, $C(p_1 \ldots p_m)$ has a $\Delta$-consistent substitution instance, $C(D_{p_1} \ldots D_{p_m})$, and so must itself be $\Delta$-consistent. □

[27] This is closely related to the problem of finding maximal superintuitionistic logics with the disjunction property. Indeed, the intuitionistic analogue of Med has this property (see Maksimova, 1986; Chagrov, 1992 and the references found therein). For a survey, including a discussion of the related properties for modal systems, see Chagrov and Zakharyashchev (1991).

## 4.7 The Substitutional Approach to Modal Predicate Logic

The present paper has focused on modal propositional logic. But it is natural also to consider the application of the substitutional approach to modal predicate logic. This raises some interesting questions, which provide a new slant on some familiar approaches to the interpretation of modal predicate logic. We shall briefly discuss these questions although they call for a far more detailed treatment.

The introduction of identity into the language brings to the forefront some difficulties for the logical interpretation of $\Box$ that are closely related to difficulties Quine raised for modal logic more generally.[28] Quine noted that certain notions, like analyticity and logical truth, are not closed under the substitution of co-referential terms. For instance, 'Hesperus' is co-referential with 'Phosphorus' yet 'Hesperus is Hesperus' is a logical truth, whilst 'Hesperus is Phosphorus' is not. This trio of claims is perfectly consistent with Leibniz's law, for the names 'Hesperus' and 'Phosphorus' do not actually appear in the latter two assertions, but are mere lexicographic components of a quotation name for a sentence. But, for just this reason, Quine argued that we cannot move up a grade of modal involvement and introduce a sentence operator that stands to the world, so to speak, as analyticity or logical truth stands to language. Kripke (1972) famously argued that we should distinguish analyticity, logical truth, and the like from metaphysical necessity and that metaphysical necessity can be expressed by a sentential operator for which it is necessary that Hesperus is indeed identical to Phosphorus. Indeed, the inconsistency of the corresponding trio of claims involving metaphysical necessity may be interpreted as a demonstration of the necessity of identicals from Leibniz's law (Kripke, 1971, p. 163).

But these considerations take on a critical significance when our goal is to introduce a notion of *logical necessity* that stands to reality as logical truth stands to language, for we would then like the metalinguistic and metaphysical notions to line up. So how might it be maintained to be *logically* necessary that Hesperus is Phosphorus? If one could speak of the logical form of a proposition—construed as a composite of individuals, relations, and so on, as opposed to sentences or Fregean thoughts—the proposition that Hesperus is Phosphorus would involve the same object on either side of the identity relation and so any proposition with the same logical form would be true. The enquoted sentence, by contrast, does not involve the same name twice: the discrepancy arises because the structure of the language used to express the propositions does not perfectly reflect the structure of the underlying reality—we have two different names for the same object. We ensured that this problem would not arise when stating the substitutional, metalogical and Tarskian constraints on logical necessity by insisting, that we only accept those instances of the relevant schemas that were formulated in a logically perfect language in Russell's sense of the term, i.e. in a language in which non-logical constants do not denote logically complex properties and no two non-logical constants co-denote. In the present section we will show, among other things, how our treatment of logical necessity might be extended to a to logically imperfect language.

[28] As, for example, in *Reference and Modality* from Quine (1953).

An *interpretation* for a first-order language (without modality) may be taken to consist of a non-empty domain $D$ of individuals and a function $\delta$ which assigns an individual from $D$ to each individual constant and a subset of $D^n$ to each $n$-place non-logical predicate letter. We shall assume, as usual, that $=$ is assigned the identity relation on $D$. We shall, for convenience, suppose that the models are *full* in the sense that for each individual $d \in D$ there is an individual constant $a$ for which $\delta(a) = d$—i.e., $\delta$ is surjective on individuals. An interpreted first-order language is *logically perfect* (with respect to the individual constants) when $\delta$ is injective.

The notion of a (concrete) substitution function and of its action on an arbitrary formula is now more complicated than in the propositional case. We should allow each individual constant to be replaced by another individual constant and we might also allow it to be replaced by a complex term should these be taken to be part of the original language (either through the use of a description operator or function letters). Similarly, we should allow each non-logical predicate letter to be replaced by another non-logical predicate letter, and we might also allow it to be replaced by any complex predicate expression. In order to indicate how a complex predicate expression is to apply to its arguments, we might indicate it by something like a $\lambda$-term $\lambda x_1 x_2 \ldots x_n . A$ and then, in making the substitution, we should take care to avoid any unintended clash of variables (a similar problem will also arise for complex individual terms should they be allowed to contain the description operator or other variable-binding operators). This means that we no longer have a simple compositional rule according to which for any substitution $i$:

$$i(Pt_1 t_2 \ldots t_n) = iP(it_1 it_2 \ldots it_n)$$

since the result of substituting $\lambda x_1 x_2 \ldots x_n A(x_1, x_2, \ldots, x_n)$, for example, for $P$ will be $A(t_1, \ldots, t_n)$ rather than $\lambda x_1 x_2 \ldots x_n A(x_1, x_2, \ldots, x_n) t_1 t_2 \ldots t_n$. We could, on the other hand, already allow complex predicate expressions to belong to the language and thereby avoid this difficulty, although at the expense of complicating the syntax of the language.

We can now bring to bear our earlier remarks concerning a Russellian logically perfect language. For the instances of McKinsey's schema—that '$\Box A$' is true iff every substitution instance of $A$ is true—will depend on the language in which it is formulated. But as we have argued, in a logically imperfect language there may be instances of the schema that are false. For consider a valuation $v$ for which $v(a = b) = 1$ and $v(a = c) = 0$ (where $a$ and $b$ are co-referential while $a$ and $c$ are not, as can happen in a logically imperfect language). Now $v(\Box(a = b)) = 1$ only if $v(i(a = b)) = 1$ for any allowable substitution $i$. But consider the substitution $i$ of $a$ for $a$ and $c$ for $b$. Then $v(i(a = b)) = v(a = c) = 0$; and so, if this substitution is allowed, $v(a = b \rightarrow \Box(a = b)) = 0$. But this violates the Necessity of Identity. It also leads to a violation of Leibniz's law, which we have suggested should be valid for any notion of propositional necessity. For substituting an arbitrary term for $a$ in $a = a$ yields a truth. Hence $v(\Box a = a) = 1$ and $v(a = b \rightarrow \Box a = a \rightarrow \Box a = b) = 0$.

Of course, no ordinary language will be logically perfect. How then might we adopt a substitutional interpretation of necessity for such a language and yet still save

Leibniz's law? One possible line of solution, considered in Fine (2005, pp. 51–52, 109–110), is to introduce a notion of logical form of a sentence that takes into account the referents of the individuals constants. Thus 'Hesperus is Phosphorus' will be taken to be of a different logical form than 'Hesperus is Mars', since the former pair of terms are coreferential while the latter pair are not; and on the resulting conception of logical truth, 'Hesperus is Phosphorus' will count as a logical truth. What this means is that in logically imperfect languages, we should impose the constraint that substitutions preserve co-reference: if $\delta(a) = \delta(b)$ then $\delta(ia) = \delta(ib)$.[29] In a logically perfect language this constraint is vacuously satisfied, since $\delta(a) = \delta(b)$ only when $a$ and $b$ are the same term; and a requirement of this sort—in which semantical elements are allowed to make a contribution to logical form—is quite different from anything that arises in the propositional case.

Even with the assumption that we are working within a logically perfect language or with the above constraint on substitutions, we may still generate counterexamples to the necessity of distinctness:

$$a \neq b \rightarrow \Box a \neq b$$

For surely we can have two names for two different objects, $a$ and $b$, whether our language is logically perfect or not. So $v(a \neq b) = 1$. Consider now the substitution $i$ of $a$ for $a$ and of $a$ for $b$. Then $v(i(a \neq b)) = v(a \neq a) = 0$, and so $v(\Box a \neq b) = 0$, assuming that $i$ is indeed an allowable substitution.

One option here is to embrace these failures of the necessity of distinctness.[30] For unlike contingently true identities, contingently true distinctness claims are consistent with the usual logical axioms governing quantification and identity. It is only in the presence of Brouwer's principle, or one of its various weakenings, that the necessity of distinctness becomes a theorem of quantified modal logic (Prior, 1967, p. 146).[31] Yet we have already considered an argument of McKinsey in Sect. 4.4.1 that Brouwer's principle is not part of the logic of logical necessity. Perhaps we should say the same

[29] This corresponds to the requirement that the interpretation of identity should be persistent in a Kripke model of quantified S4. More precisely, we may define a Kripke model $(W, R, V)$ just as in Proposition 4.2.11 in which $W$ is the set of substitutions, $R = \{(i, j \circ i) \mid i, j \in W\}$ and $V(i, Ra_1 \ldots a_n) = v(i(Ra_1 \ldots a_n))$ when $R$ is any relation or predicate symbol (including identity). If $V(i, a = b) = 1$, then $v(i(a = b)) = 1$ so $\delta(ia) = \delta(ib)$, and since all substitutions preserve co-reference, $\delta(jia) = \delta(jib)$ for any $j \in W$. So $v(ji(a = b)) = V(j \circ i, a = b) = 1$, which means that identity is persistent.

[30] See Bacon (2020, pp. 554–556).

[31] The proof in Prior (1967) is attributed to E.J. Lemmon. An earlier proof of the necessity of distinctness within a the stronger background logic of S5 may be found in Prior (1955, p. 206), concerning which Prior writes 'the obscurity of the issue does leave S5 under a measure of suspicion, and this must be set against our earlier argument from simplicity in its favour' (p. 207). (The necessity of distinctness may also be derived from various weakenings of Brouwer's principle, such as $A \rightarrow \Box\Diamond\Diamond A$, or $\Box(A \rightarrow \Box A) \rightarrow \Diamond A \rightarrow A$. However, in the present substitutional setting in which the principles of S4 are being granted, these variant principles are *not* strict weakenings of Brouwer's principle but equivalent to it.)

about the necessity of distinctness? (For there is, after all, a striking resemblance between McKinsey's argument and the above argument against the necessity of distinctness.)

We could, alternatively, ensure the necessity of distinctness by imposing additional constraints on substitutions. For just as we previously required that co-referential constants be substituted for co-referential constants, we could now require that non-co-referential constants be substituted for non-co-referential constants.[32] For we might wish to insist, at least in the case of the individual constants, that a substitution should preserve the exact logical form of the original expression—distinctions of logical form within the original expression should not be made to disappear. Thus, in the present case, $a$ cannot be substituted for both $a$ and $b$ since $a = b$ and $a = a$ are not of exactly the same logical form. And again, this is a requirement that has no counterpart—or, at least, no significant counterpart—in the propositional case.

We might wish to further restrict or modify the constraints on which substitutions on individual constants should be allowed. Let $IC$ the set of individual constants and $PL$ the set of predicate letters of the language in question. We may then represent a substitution as a partial function from $IC \cup PL$ subject to the requirement that individual constants go into individual terms and predicate letters go into predicate expressions. We may then suppose that we are given an arbitrary set $S$ of allowable substitutions subject only to the requirement that the restriction of any substitution in $S$ is also a substitution $S$. Since the functions in $S$ are now partial, the clause for $\Box A$ should be modified accordingly, so that $v(\Box A) = 1$ iff $v(iA) = 1$ for all allowable substitutions $i$ defined on $A$.

Further conditions of a general structural nature might then be imposed on $S$. It might be required, for example, that each partial function in $S$ should be extendible to a total function or that the functions on individual constants and on predicate letters should be freely combinable, so that if $i$ and $j$ are two partial functions in $S$ then the partial function $i \restriction IC \cup j \restriction PL$ should also be in $S$. But the former requirement might be questioned if it was thought that the individuals designated by the constants in the range of a substitution function should somehow be compossible; and even the latter requirement might be questioned if it was thought that a predicate letter $P$ could contain a "hidden" individual constant $a$, so that a restriction on what could be substituted for a would result in a restriction on what could be substituted for $P$.

Constraints of a more particular character might also be imposed on $S$. It might be supposed, for example, that certain constants $a$ (including the special constants that correspond to an individual a) should be "rigid", so that only $a$ can be substituted for $a$. This would provide us with another way of securing the validity of the principle $a = b \rightarrow \Box(a = b)$ and of the principle $a \neq b \rightarrow \Box a \neq b$ in the case in which $a$ and $b$ are rigid constants, since the allowable substitutions would leave them intact. It would also allow us, given that certain non-logical predicates were allowed to be rigid, to sustain the truth of various substantive essentialist claims. If, for example, 'Felix' was a rigid constant for a tiger and 'tiger' a rigid predicate, then Felix would

[32] This corresponds to the requirement that the interpretation of identity be anti-persistent in the Kripke model associated with a valuation $v$, i.e., if $j \circ i \Vdash a = b$ then $i \Vdash a = b$. See Footnote 28.

necessarily be a tiger ($v(\Box Tiger(Felix)) = 1$) while it might well be false that Felix is necessarily in a zoo (($v(\Box Zoo(Felix)) = 0$) given that the predicate 'Zoo' is not rigid, since a predicate $P$ for which $P(Felix)$ is false might then be substituted for 'Zoo'. In this way, we could, to some extent, mimic an essential view of individual objects within the substitutional framework.

We could, in a similar way, mimic Lewis's counterpart theory. For when Lewis would say that $a'$ is a counterpart of $a$, we might say that a′ can be substituted for a or when one might say, more generally, that $a'$, $b'$, $c'$, … are relational counterparts of $a$, $b$, $c$, …, we might say that a′, b′, c′, … can be simultaneously substituted for a, b, c, …. Our substitution-theoretic clause for necessity would then correspond to his counterpart-theoretic clause. We thereby obtain a highly non-realist version of counterpart theory, one that might be especially suited to someone who wished to adopt a linguistic view of necessity. We should also note that, just as we might wish the substituents for predicate letters with "hidden" individual constants to vary with those for the individual constants, so one might wish the counterparts for the non-qualitative properties expressed by predicate letters to vary with those for the individuals designated by individual constants, even though this was no part of Lewis' original proposal.[33]

Quantifiers give rise to further issues. Within the present setting, it is natural to adopt a substitutional interpretation of the quantifier. Thus we can either say:

$v(\forall x A(x)) = 1$ iff $v(A(t)) = 1$ for all terms $t$

thereby obtaining something akin to a "conceptual" interpretation of the quantifier, or we can say:

$v(\forall x A(x)) = 1$ iff $v(A(t)) = 1$ for all rigid terms $t$,

thereby obtaining something akin to an "objectual" interpretation of the quantifiers. Let us restrict our attention to the objectual interpretation although some related issues also arise for the conceptual interpretation.

Then under the innocuous assumption that a rigid constant can always be substituted for itself, we can establish the equivalence of $\forall x \Box A(x)$ to $\Box \forall x A(x)$ ($v(\forall x \Box A(x)) = v(\Box \forall x A(x)$ for any valuation $v$). How might this equivalence be avoided, i.e., how might we gain the effect of a variable domain of quantification? One possibility is to suppose that our quantifiers are restricted. Thus universal quantifications now take the form $\forall x(A(x) : B(x))$ (all A's are B's) and we set $v(\forall x(A(x) : B(x))) = 1$ iff for every rigid term a either $v(A(a)) = 0$ or $v(A(a)) = 1$.

We now take the unrestricted quantification $\forall x A(x)$ to be implicitly restricted to a dummy domain predicate letter $D$. Thus $\forall x A(x)$ is read as $\forall x(Dx : B(x))$. This will then give us the effect of a variation in the domain, since different substitutions of a predicate $P$ for $D$ will provide us with different range of rigid constants a for

[33] Dorr (2005) discusses a related difficulty which arises when the language allows us to quantify over propositions but, if we are right, the difficulty can already be seen to arise within the original framework of modal first-order predicate logic.

which $v(Pa) = 1$. The equivalence of $\forall x \Box A(x)$ to $\Box \forall x A(x)$ will then fail in both directions. For $\forall x \Box A(x)$ will be read as $\forall x(Dx \rightarrow \Box A(x))$ while $\Box \forall x A(x)$ will be read as $\Box \forall x(Dx \rightarrow A(x))$; and as long as the domain predicate $D$ is not rigid, neither formula will be equivalent to the other under a constant domain interpretation of the unrestricted quantifier.

As should be clear from this brief discussion, the substitutional approach has enormous flexibility in its application to modal predicate logic. It is able to accommodate a wide variety of interpretative stances—necessary versus contingent identity, objectual versus conceptual quantification, rigid versus counterpart-theoretic accounts of de re modality, fixed versus variable domains; and it may also be able to provide a certain degree of succor for those who favor a less metaphysical understanding of de re modality.

**Acknowledgements** We pay cross-generation tribute to Saul's work on modal logic, one of us having learned of his work at its outset and the other having learned of it after several decades of development. He has done more than anyone to put the semantical study of modal logic on a firm technical footing, and our debt to his pioneering work should be evident throughout the paper.
Thanks are due to Lloyd Humberstone and Peter Fritz for correspondence with AB on this material that greatly benefited the paper. Theorems 4.3.12, 4.5.6 and 4.6.1 are due to AB, and Theorems 4.3.8, 4.5.8, 4.6.17 and Sect. 4.6 are due to KF. The results in Sect. 4.4.2 were joint. Other minor propositions and observations are due to AB unless otherwise indicated.

## References

Anderson, A. R., & Belnap, N. D. (1975). *Entailment: The logic of relevance and neccessity* (Vol. I). Princeton University Press.
Bacon, A. (2019). Substitution structures. *Journal of Philosophical Logic, 48*(6), 1017–1075. https://doi.org/10.1007/s10992-019-09505-z
Bacon, A. (2020). Logical combinatorialism. *Philosophical Review, 129*(4), 537–589. https://doi.org/10.1215/00318108-8540944
Bernays, P., & Schönfinkel, M. (1928). Zum entscheidungsproblem der mathematischen logik. *Mathematische Annalen, 99*(1), 342–372.
Carnap, R. (1946). Modalities and quantification. *Journal of Symbolic Logic, 11*(2), 33–64. https://doi.org/10.2307/2268610
Carnap, R. (1947). *Meaning and necessity*. University of Chicago Press.
Carroll, M. J. (1978). An axiomatization of S13. *Philosophia, 8*(2–3), 381–382. https://doi.org/10.1007/BF02379250
Chagrov, A., & Zakharyashchev, M. (1991). The disjunction property of intermediate propositional logics. *Studia Logica, 50*(2), 189–216. https://doi.org/10.1007/BF00370182
Chagrov, A. V. (1992). Continuality of the set of maximal superintuitionistic logics with the disjunction property. *Mathematical Notes, 51*(2), 188–193.
Cocchiarella, N. B. (1974). Logical atomism and modal logic. *Philosophia, 4*(1), 41–66. https://doi.org/10.1007/bf02381515
Cresswell, M. (2013). Carnap and Mckinsey: Topics in the pre-history of possible-worlds semantics. In *Proceedings of the 12th Asian Logic Conference* (pp. 53–75).
Cresswell, M. J., & Hughes, G. E. (1996). *A new introduction to modal logic*. Routledge.

Dorr, C. (2005). Propositions and counterpart theory. *Analysis, 65*(3), 210–218. https://doi.org/10.1111/j.1467-8284.2005.00551.x
Drake, F. R. (1962). On Mckinsey's syntatical characterizations of systems of modal logic. *Journal of Symbolic Logic, 27*(4), 400–406. https://doi.org/10.2307/2964546
Fine, K. (n.d.). *Modal logic as metalogic*. (Unpublished.)
Fine, K. (1970). Propositional quantifiers in modal logic. *Theoria, 36*(3), 336–346. https://doi.org/10.1111/j.1755-2567.1970.tb00432.x
Fine, K. (1977). Properties, propositions and sets. *Journal of Philosophical Logic, 6*(1), 135–191. https://doi.org/10.1007/bf00262054
Fine, K. (2005). *Modality and tense: Philosophical papers*. Oxford University Press.
Friedman, H. (1975). One hundred and two problems in mathematical logic. *Journal of Symbolic Logic, 40*(2), 113–129. https://doi.org/10.2307/2271891
Fritz, P. (2024). *Propositional quantifiers* (1st ed.). Cambridge University Press.
Hamkins, J. D., Leibman, G., & Löwe, B. (2015). Structural connections between a forcing class and its modal logic. *Israel Journal of Mathematics, 201*, 617–651.
Hendry, H. E., & Pokriefka, M. L. (1985). Carnapian extensions of S5. *Journal of Philosophical Logic, 14*(2), 111–128. https://doi.org/10.1007/BF00245990
Hilbert, D., & Ackermann, W. (1928). *Grundzüge der theoretischen logik*. Berlin: Julius Springer.
Holliday, W. (2017). On the modal logic of subset and superset: Tense logic over Medvedev frames. *Studia Logica, 105*(1), 13–35. https://doi.org/10.1007/s11225-016-9680-1
Humberstone, L. (2016). *Philosophical applications of modal logic*. College Publications.
Jerabek, E. (2023). An extension of the disjunction property in modal logic. MathOverflow. https://mathoverflow.net/q/453410
Kripke, S. (1972). *Naming and necessity*. Harvard University Press.
Kripke, S. A. (1959). A completeness theorem in modal logic. *Journal of Symbolic Logic, 24*(1), 1–14. https://doi.org/10.2307/2964568
Kripke, S. A. (1971). Identity and necessity. In M. K. Munitz (Ed.), *Identity and individuation* (pp. 135–164). New York: New York University Press.
Kripke, S. A. (1976). Is there a problem about substitutional quantification? In G. Evans & J. McDowell (Eds.), *Truth and meaning truth and meaning* (pp. 324–419). Oxford University Press.
Maksimova, L. L. (1986). On maximal intermediate logics with the disjunction property. *Studia Logica, 45*(1), 69–75.
McKinsey, J. C. C. (1945). On the syntactical construction of systems of modal logic. *Journal of Symbolic Logic, 10*(3), 83–94. https://doi.org/10.2307/2267027
Meyer, R. K. (1971). On coherence in modal logics. *Logique Et Analyse, 14*, 658–668.
Prior, A. (1955). *Formal logic*. Oxford University Press.
Prior, A. (1967). *Past, present and future*. Clarendon Press.
Prucnal, T. (1979). On two problems of Harvey Friedman. *Studia Logica, 38*(3), 247–262. https://doi.org/10.1007/BF00405383
Quine, W. V. O. (1953). *From a logical point of view*. Harvard University Press.
Russell, B. (1940). *The philosophy of logical atomism*. Routledge.
Thomason, S. K. (1973). A new representation of S5. *Notre Dame Journal of Formal Logic, 14*(2), 281–284. https://doi.org/10.1305/ndjfl/1093890907
Urquhart, A. (2010). Anderson and Belnap's Invitation to Sin. *Journal of Philosophical Logic, 39*, 453–472.
Williamson, T. (2013). *Modal logic as metaphysics*. Oxford University Press.

# Chapter 5
# Relational Patterns, Partiality, and Set Lifting in Modal Semantics

**Johan van Benthem**

**Abstract** We articulate a relational understanding of modality, and show it at work in a survey of pre-order models, set-lifting, and modal languages.

**Keywords** Accessibility relation · Pre-order model · Language design · Set lifting · Logic translation

## 5.1 Modality as a Relational Pattern

People often demand to know from me "what a possible world is".[1] How can they understand modal logic, if I do not tell them this basic simple thing first? When I hear such questions, alarm bells go off in my mind.

**What and how questions**. As a student, I was raised in the spirit of Evert Beth, an erudite philosopher and creative logician. His magnum opus is "Foundations of Mathematical Logic" (Beth, 1959), and a chapter that impressed me is that on 'Aristotle's Theory of Science' with its focus on "What Is" questions. For Beth, this agenda became obsolete with the rise of modern science, where the deep insights came from asking "How" questions about how Nature functions. Beth felt that, despite the efforts of Russell and his generation in bringing together science and philosophy,

[1] In the 1960s, Saul Kripke opened the doors to modern modal logic, and I was one of a generation of students entering this wonderland of philosophical ideas pursued with mathematical clarity. Over the years, many further influences have shaped the field, coming also from linguistics and computer science, and I am a representative of the resulting 'Amsterdam style' in modal logic, cf. Blackburn et al. (2001), van Benthem et al. (2006). But whatever agenda one lives by, one constant in the field is our shared indebtedness to Saul. It is a pleasure to make this small offering in his honor.

J. van Benthem (✉)
ILLC, University of Amsterdam, Amsterdam, The Netherlands
e-mail: j.vanbenthem@uva.nl

Department of Philosophy, Stanford University, Stanford, CA, USA

Joint Research Center for Logic, Tsinghua University, Beijing, China

© The Author(s), under exclusive license to Springer Nature Switzerland AG 2024

Y. Weiss and R. Birman (eds.), *Saul Kripke on Modal Logic*, Outstanding Contributions to Logic 30, https://doi.org/10.1007/978-3-031-57635-5_5

Aristotle's essentialist mindset still held sway widely. However this may be, ever since my student days, I have felt attracted to a philosophical "How" approach, as in David Lewis' famous saying that to know what a meaning *is*, one must know what a meaning *does* (Lewis, 1970).

**Abstraction**. To me, possible worlds semantics is a Carnapian conceptual framework, and worlds are objects in abstract modal structures. When you use them to represent things, they can be just anything you like: points in time, situations, deals of cards, people. Abstraction is a cognitive talent and a driver in intellectual history. Being able to 'let go' of pre-set interpretations is an intellectual talent that fosters understanding at the level where it thrives best.[2]

**An interactive relational perspective**. Still, one can say more about worlds, even at an abstract level. Consider the relational models $\mathbf{M} = (W, R, V)$ introduced by Kanger and Kripke. Critics see the accessibility relations $R$ between worlds as an ad-hoc device for modeling some logics that would lack completeness theorems otherwise. My own view is the opposite: the relational structure is the heart of the matter. A possible world may have some internal structure of local facts, but its truly modal behavior is its *pattern of interaction* with other worlds. To paraphrase Zellig Harris' famous dictum about word meanings, "to understand a world, look at the company it keeps". This reflects the message of modern logic promoted by Russell and his generation. Objects are not just monads determined by their properties; equally important are the binary and other relations encoding their relationships to other objects: hence the change from the traditional Aristotelean logic of properties to a Fregean logic of predicates.[3]

I would go even further, and say that the modal content of a model *is* its relational pattern: modality lives 'in the air', not inside the worlds—it is team play. These modal patterns arise in many settings, leading to such different families as modal logics of information and knowledge, time and action, space, metaphysics, and more as the scope of modal logic keeps growing.

**Invariant patterns and modal languages**. But the relational perspective, once taken, is pervasive. In particular, a modal model is just one representative of the relational 'pattern' that I mentioned. The real structural pattern is the equivalence class of that model under the appropriate structural *invariance relation*. There are many fundamental invariance relations in mathematics and science, and logic follows suit, with

[2] One can of course legitimately criticize an abstract semantics for fit to some concrete application. Such 'criticisms of fit' are in fact one engine of progress in the field.

[3] Saying that worlds are determined by how they relate to other worlds may sound circular. But think of Category Theory, where categories describe objects and morphisms, not through internal structure, but in a pattern of external interactions. Since Kripke's pre-order models for intuitionistic logic and modal S4 are categories, this analogy is perfectly felicitous, and modal logic is mathematically sound. However, I admit there may be legitimate metaphysical worries. If an object is determined by its interactions with other objects, changing its context will automatically change the object: this weathervane behavior may be considered problematic.

relations such as isomorphism, homomorphism, or for basic modal logic: bisimulation (Blackburn et al., 2001).[4] This takes us to a next aggregation level: possible worlds models do not live alone, but in a modal universe of models, and the identity of a model is also determined by its relations to those other models.

In this setting, *modal languages* enter, poorer or richer, offering perspicuous ways of expressing invariant relational structure of models. There is an extensive literature on pattern invariance and definability in matching modal languages (Goranko & Otto, 2006). Thus, there is an equally important syntactic side to modal logic, which consists in the study of different languages for describing different relational patterns up to different levels of expressive power, depending on how structurally discerning we make the invariance relation.

These modal languages offer a generic perspective going beyond peculiarities of specific models whose complete structure may be infinitely complex. We can describe given models up to some level of syntactic complexity, using normal forms, modal Scott sentences, or even the drastic information reduction of the modal method of filtration (van Benthem & Bezhanishvili, 2021). I will not emphasize this important generic function in what follows, but fine-structure and genericity is what all languages, logical or natural, are for.[5] Also, modal syntax reveals characteristic patterns of reasoning of wide application, one more strength of the framework that will fall outside the scope of this paper.

**Modal logics**. Finally, different modal languages induce different logical systems of valid inference forms, often axiomatizable in modal proof systems.[6] Of course, our relational perspective also applies to these languages and logics. Much about their identity is revealed through *translations* into their neighbors: other languages and logics, and this, too, will be a theme in what follows.[7]

My relational context-oriented stance at every level may seem obsessive, but to me, awareness of this modal structure is really turtles all the way down.[8] Even so, I admit that the relational view can be challenged. There are also forces inside modal logic that locate the essence in a monad-style view of worlds. The second Appendix

[4] The definition of bisimulation (Blackburn et al., 2001) reflects what we said above about worlds: matched worlds only need to agree on a few thin atomic properties; the modal structure is in the back-and-forth conditions on available relational steps.

[5] One may wonder what makes a *language* modal, as opposed to any language for relational structures, say first- or second-order. Common answers are decidable validity or perspicuous variable-free notation for proof theory, but the border-line between modal and just any logical languages remains a bit elusive. In fact, switching perspectives is often helpful, with the 'standard translation' from modal into first-order languages as a prime example.

[6] The above expressive diversity is not the usual deductive diversity of different 'modal logics' over the same language, say: K, S4, or S5. Deductive diversity is best understood as reflecting different restrictions one can make on the relational patterns one is interested in.

[7] For more on modal proof theory as a subfield with its own intuitions and traditions, the reader is referred to Wansing (1996), Fitting (2007), Negri (2011).

[8] If you see *Coalgebra* (Jacobs, 2016) as the fresh young face of modal logic, then turtles all the way down, not bottom-up construction, is the non-wellfounded essence of modality.

contains further thoughts on the role of modal languages, and on design of modal languages for all the three relational levels mentioned here.

**Content of this paper**. The preceding themes become more concrete with some history of ideas. In the course of this paper, I will discuss partial possibilities, set-lifting and language design, since these pervasive topics relate to recent challenges to modal logic coming from situation theory, truthmaker semantics, or hyper-intensionality generally. As we shall see, these contemporary challenges have an interesting past, and their implementation involves doing more, rather than less, modal logic, requiring richer relational patterns than the sparse models of the founding generation. The presentation to follow is a light historical survey with commentary, and it can hopefully be appreciated even if one does not share, or rejects, my patternist stance.

## 5.2 Intuitionistic Logic and Modal S4

Let us start with the classic paper "Semantic Considerations on Intuitionistic Logic I" (Kripke, 1965). Our further themes will unfold from there.[9]

**Basic language and semantics**. Kripke models for the language of intuitionistic propositional logic are triples $\mathbf{M} = (W, \leq, V)$ where $(W, \leq)$ is a reflexive transitive pre-order, and the valuation $V$ assigns sets of points in $W$ that are *upward closed*: if $s \in V(p)$ and $s \leq t$, then $t \in V(p)$. One can also restrict attention to *partial orders*, but this is not essential to what follows. The truth definition explains when a formula $\varphi$ is true at a point $s$ in a model $\mathbf{M}$ (written $\mathbf{M}, s \models \varphi$), with the following central recursive clauses:

$\mathbf{M}, s \models \varphi \wedge \psi$ iff $\mathbf{M}, s \models \varphi$ and $\mathbf{M}, s \models \psi$

$\mathbf{M}, s \models \varphi \vee \psi$ iff $\mathbf{M}, s \models \varphi$ or $\mathbf{M}, s \models \psi$

$\mathbf{M}, s \models \neg\varphi$ iff for no $t \geq s$, $\mathbf{M}, t \models \varphi$

$\mathbf{M}, s \models \varphi \rightarrow \psi$ iff for all $t \geq s$, if $\mathbf{M}, t \models \varphi$, then $\mathbf{M}, t \models \psi$

With this understanding of the logical operators, all formulas denote upward-closed sets of points. This semantics fits intuitionistic propositional logic: a formula is universally valid iff it is provable in Heyting's standard proof system.

**Geometry of information and inquiry**. While the completeness theorem shows 'fit' of validity and provability,[10] a broader virtue in this analysis is the *conceptual*

[9] Moreover, but this almost goes without saying, the equally classic paper "Semantic Considerations on Modal Logic" (Kripke, 1963) lies at the basis of the explicit modal perspectives that will be brought to bear throughout this paper.

[10] Tighter connections between semantic structures and proofs themselves are the 'full completeness' results in more category-theoretic frameworks (Abramsky & Jagadeesan, 1994).

*framework* as a style of thinking. We can now analyze modal axioms or proof principles in terms of an easily visualized 'geometry of information', or, viewing models as possible histories of investigation, the 'geometry of inquiry'.[11]

**The package**. Even so, it is good to step back and reflect a bit. Many decisions went into making intuitionistic logic the system of reasoning here. There was a choice of language (the vocabulary of propositional logic), type of models, constraints put on the ordering (reflexivity and transitivity), special properties demanded of propositions (upward-closure), and specific truth conditions for the logical operations linking language and models. Like in many modal semantics, this is a finely tuned conceptual machinery with lots of interlocking moving parts that could easily have been set differently. For instance, classical logic emerges by changing the clause for disjunction from immediate to 'cofinal choice', and requiring denotations $X$ of formulas to satisfy upward-closure plus the property of *regularity*: "if $\forall t \geq s \, \exists u \geq t : u \in X$, then $s \in X$".[12]

**Structures**. Though a very simple mathematical structure, around 1960, preorders were the right choice at the time. In fact, they are special cases of the topological semantics for intuitionistic and modal logics which had been known since the 1930s,[13] but specialization was progress. Preorders are also close in spirit to Beth tree models for intuitionistic logic (Beth, 1956), but their simplicity won the day.[14] In fact, far beyond intuitionistic logic, Kripke's models offer a setting where many logics can be defined and compared, as we shall see.

**Two states in the same land: the modal logic S4**. Intuitionistic logic famously shares the land of pre-orders with classical propositional modal logic, where the induced logic is S4. With the latter, the style of thinking is different. The classical understanding of the Boolean operations is retained, but the new order structure $\leq$ is described by matching universal and existential modalities, making S4 an 'explicit' classical counterpart to the 'implicit' non-classical intuitionistic logic on the same models (van Benthem, 2019).

Comparisons between the two neighbors can run in various ways. In terms of structural invariance relations, the language of modal S4 has a natural and well-known fit with the ubiquitous notion of bisimulation, whereas an invariance analysis

[11] For a concrete example, consider the *Disjunction Principle* for intuitionistic logic: if $\varphi \vee \psi$ is provable, then so is $\varphi$ or $\psi$. A semantic proof for this takes pointed countermodels for $\varphi$ and $\psi$, forms their disjoint union, and puts the distinguished points under a new root: in the resulting model, the disjunction is refuted. In fact, the semantic content of the Disjunction Principle is *precisely* the model glueing property of joint rooting, but a detailed correspondence perspective on modal derivation rules would be a subject for a more technical paper.

[12] In topological terms, this is a restriction to the Boolean algebra of regular open sets. One can also think of a Sure Thing Principle: "if you see that $\varphi$ is inevitable, accept it right now".

[13] Preorders match special 'Alexandroff topologies' (van Benthem & Bezhanishvili, 2007).

[14] Of course, these issues are never settled once and for all. Dummett and van Dalen have argued that Beth models are superior for some foundational purposes, and an attractive mathematical perspective taking them on board is found in Bezhanishvili and Holliday (2019).

of intuitionistic logic requires more care, involving entangled directed simulation relations (Olkhovikov, 2013; Paterson, 1997).

**Translations**. Surely the most famous connection between the two great pre-order logics is one of syntactic *translation*. Gödel (1933) translated intuitionistic propositional logic faithfully into modal S4, and Kripke's truth definition presented above is a semantic transposition of this recursive translation. The translation might be seen as a case of embedding a poorer system into a richer one, since modal S4 does not just govern reasoning about upward-persistent formulas $\Box\varphi$, but also about non-persistent formulas $\Diamond\varphi$. Yet a surprise was in store. Answering a question of Dana Scott, Fernández Duque (2006) gave a converse translation, embedding S4-validity faithfully into intuitionistic validity—with an important correction in Gore and Thomson (2019).[15]

If one truly knows a logic through its neighbors, as we have suggested, all this shows how relations between logics can be delicate—especially, given the lack of a systematic theory of translatability (and not to forget: non-translatability) between logical languages and systems. We do not know all that much about this modal meta-realm.[16,17] These considerations point to a much larger issue of system identity, and when we would call two logics 'the same': van Benthem (2019) has some discussion, but in general, these are large unresolved issues.

**Conclusion**. Pre-orders with an intuitive reading as information states or histories of inquiry offer a rich playground for modeling and comparing existing logics. This marriage is a happy one: in modal logic since the 1960s, intuitionistic and classical logics have often been studied in tandem, to mutual benefit. In the following sections, we will expand this co-existence to more logics.

One can also view this section as adding *fine-structure* to classical models. Points in an intuitionistic model are stages toward a complete classical valuation, persistence holds for assertions that are definitely true, while modal S4 adds assertions about possibilities that may be present for a while and then drop out.[18] The inclusion order embodies this fine-structure of partial approximation, and what its presence leads to is more, rather than less modal logic.

---

[15] A related earlier translation in Friedman and Flagg (1986) was only sound, not faithful.

[16] In particular, translations show that appearances qua presentation may be misleading. Prima facie, intuitionistic propositional logic is a sublogic of classical logic, but in reality, it is provably more complex in terms of validity and proof search, having a Pspace-complete satisfiability problem as opposed to the NP-completeness for classical propositional logic.

[17] Two recent dissertations discussing the landscape of translations between modal logics are French (2011), Kocurek (2018). General model-theoretic perspectives on relative interpretability can be found in van Benthem and Pearce (1984), Barrett and Halvorson (2016).

[18] This is like in formal learning theory (Kelly, 1996), where we only know the complete truth in a completed history, while knowledge and belief play on the way. Note that this form of knowledge is *procedural*: it refers to a current process of inquiry, and thus differs from more standard static conceptions in epistemic logic of knowledge as a range of options. van Benthem (2009) discusses Kripke's models extensively from both epistemic perspectives.

## 5.3 Languages and Logics on Information Models

**Semantics or language design**. Once on the level playing field of pre-orders, further aspects of information and inquiry can be studied. This represents a shift in direction from what is suggested by the term 'possible worlds *semantics*'. A semantics is usually given for some existing language, which sets the standards for what proposed models are meant to achieve. In this sense, we have discussed semantics for an intuitionistic language and its associated reasoning practice, or for a language with modal expressions of, say, knowledge or necessity. But one can also take proposed structures as primary, and ask what sort of language would best bring out important features of reasoning in this setting. Both directions occur in modal logic, with new language design perhaps more prominent in logics of time and action. We will continue in the latter mode, but in our story, the two directions are not mutually exclusive or opposed.

*Note*: Henceforth, we will drop the term 'world' and use the neutral 'point'.

**A richer modal information logic**. Our first offering is a system proposed in van Benthem (1996). One natural structure in pre-orders are *suprema* (lowest upper bounds) and *infima* (greatest lower bounds) of states. On partial orders, existing suprema and infima will be unique, but one need not assume this in general. Here, a supremum of points $s$ and $t$ can be seen as modeling a 'merge' of two information states, or alternatively, as a meeting point where two lines of inquiry come together. This suggests adding two binary modalities $\langle \text{sup} \rangle \varphi$ and $\langle \text{inf} \rangle \varphi$ to the basic modal language, with the following interpretation:

$\mathbf{M}, s \models \langle \text{sup} \rangle \varphi \, \psi$ iff there exist two points $t, u$ such that

(a) $s$ is a supremum of $t, u$, (b) $\mathbf{M}, t \models \varphi$ and (c) $\mathbf{M}, u \models \psi$.

and likewise for $\langle inf \rangle \varphi \psi$. In this language, we can define the standard modalities in both directions. For instance, $\langle \leq \rangle \varphi$ is equivalent to $\langle inf \rangle \varphi \top$.

The new binary modalities satisfy counterparts of familiar modal principles, including distribution over disjunctions, giving the system a natural feel.

As for further principles, we did not assume that suprema and infima always exist in our models. Doing so would create a semi-lattice or lattice structure that validates new principles beyond the minimal logic. However, these can still be analyzed by modal frame correspondence techniques, as many principles in this setting have so-called 'Sahlqvist syntax' (Blackburn et al., 2001). An example of an additional principle is Associativity

$$\langle \text{sup} \rangle (\langle \text{sup} \rangle \varphi \, \psi) \, \alpha \rightarrow \langle \text{sup} \rangle \varphi (\langle \text{sup} \rangle \psi \, \alpha).$$

This enforces existence of enough suprema in the model to reorder arbitrary combinations of the form $sup(sup(s, t), u)$ to $sup(s, sup(t, u))$. While this seems a harmless requirement that helps computation, such existential assumptions can increase the complexity of validity: some basic modal logics of associative binary modalities are undecidable (Kurucz et al., 1995).

Perhaps surprisingly, not much is known about this simple modal information logic mildly extending the familiar S4. For instance, no complete axiomatization has been found yet. In the first Appendix, we give an axiomatization using an extra device: nominals that name specific points. This will show at least how one can work in practice with a modal language like the one presented here.[19]

Another and perhaps even more striking open problem about our simple extension of S4 is the following: *Is basic modal information logic decidable*?

**Implication, categorial logic, and perspective shifts**. Viewing points in our pre-order models as information pieces suggests further modalities beyond the above, such as the following natural *implication*, stating that combining with any evidence for the antecedent yields evidence for the consequent:

$\mathbf{M}, s \models \varphi \Rightarrow \psi$ iff for each point $u$ with $\mathbf{M}, u \models \varphi$ and each point $t$ that is a supremum of $s$ and $u$, it holds that $\mathbf{M}, t \models \psi$.

Now we are at a crossroads between traditions. This notion is reminiscent of the Brouwer-Heyting-Kolmogorov proof interpretation of intuitionistic implication (Troelstra & van Dalen, 1988), and in a more general category-theoretic setting, it satisfies the basic laws of categorial implication as codified in the commutative Lambek Calculus (Kurtonina & Moortgat, 2009). But also in classical modal logic terms, the above can be seen as a familiar device. In the presence of Boolean operators, the above implication amounts to introducing a binary modality for the ternary supremum relation $z = sup(x, y)$ as seen from the perspective of the point $x$, rather than the point $z$ used by the sum modality $\langle \text{sup} \rangle$. Language enrichments for perspective shifts in abstract ternary relation models lead to a more elegant overall design of modal systems (Venema, 1992).

**Relevant logic, information and computation**. Most strikingly, the preceding themes connect with *Relevant Logic,* which offers a general account of information structure (Bimbo & Dunn, 2008) with ternary relational models as an illuminating abstract backdrop to our modal notions on binary pre-orders. Indeed, on semi-lattices, the above notion is the relevant implication of Urquhart (1972), and our question about modal information logic may echo the still open problem whether Urquhart's positive semi-lattice logic **S** is decidable (Urquhart, 2016). As with the other paradigms featuring in this article, relevant logic contains a wide range of notions and results beyond the themes in our presentation, for which we refer the reader to Anderson et al. (2017). The latter book also links up with situation theory, one more paradigm putting information at central stage in logic (Barwise, 1989; Barwise & Seligman, 1995).[20]

[19] In that same appendix, we will also briefly discuss modal axiomatizations of the above-mentioned semi-lattices and lattices, as an instance of 'lifting' equational theories of algebraic structures to modal logics—in the spirit of Sect. 5.5 below.

[20] In a full-scale version of this paper, Relevant Logic would occupy a more prominent place, and it could then be explored how our general modal themes might play in that tradition. Cf. Weiss (2021) for a new arithmetical model that might provide a concrete case study.

On a final outreach track, related structures have come up in logics of computation. The above implication is the 'magic wand' of separation logic (Reynolds, 2002), a recent theory of data structures with memory maintenance, with a precursor in the 'logic of bunched implications' (O'Hearn & Pym, 1999).

**Updates and dynamic information modalities**. Computing is a process that generates information, and the preceding connection suggests a further dynamic perspective on our pre-order models. We follow the line of thinking in van Benthem (1989) who proposed generalizing from Brouwer's 'idealized mathematician' to ordinary cognitive agents. Moving upward along the ordering $\leq$, such agents arrive at more information about the actual world, or moving downward, they backtrack and give up things they thought they knew.

These informational movements correspond to natural *epistemic actions*. Updates with new information $\varphi$ take us from a current point to points higher up in the $\leq$ order where $\varphi$ is true, perhaps, according to a common intuition about updating, to *closest* points where $\varphi$ is true. Conversely, what may be called downdates take us to closest points lower in the order where $\varphi$ is false.

The dynamic perspective of information update in pre-order models at once suggests the introduction of a matching new modality to the usual vocabulary:

$\mathbf{M}, s \models [+\varphi]\psi$ iff for each point $t \geq s$ that is closest to $s$ among the points satisfying $\varphi$ in the model, $\mathbf{M}, t \models \psi$.

Downdate modalities are defined in a completely analogous manner referring to closest points $T$ below $s$ in the information ordering $\leq$.[21]

The resulting core logic of information update and retraction is comparable in its scope with belief revision theory (Gardenfors & Rott, 1995), or dynamic-epistemic logic (Baltag & Renne, 2016). Its update modalities also connect with *conditional logic*, being a closeness-style (Lewis, 1973; Stalnaker, 1969) variant of intuitionistic implication. With this natural simple modal system, we are again in uncharted territory, except for some forays in van Benthem (2018). The complete update logic of pre-orders remains to be determined.[22]

**Structures and invariances**. Our examples raise the question what structures are most natural. Should we specialize from pre-orders to *lattices*, where suprema and infima always exist, or to distributive lattices? Computer science has a long tradition in relevant abstract models of information. In particular, Scott's *domain theory* (Abramsky & Jung, 1994) offers a rich account of information structures as DCPOs:

[21] A setting looking both forward and backward in the ordering also occurs in bi-intuitionistic logics, where 'co-implication' refers to the past (Gore, 2000; Moisil, 1942; Rauszer, 1977).

[22] The relevant logic connection may again enter here: cf. Aucher (2014) on analyzing current dynamic-epistemic logics of information update as ternary relation-based substructural logics.

partial orders where all suprema of directed subsets exist, over which abstract computation can take place.[23]

In conjunction with this, there is also the issue of which structural invariance relations match the extended languages introduced just now. For a modal information theory over lattices, a bisimulation analogue of 'Scott continuity' might be the answer. For our conditional logic of updates and downdates, some adaptation of known bisimulations for conditional logic may be appropriate.

**Digression: partial logic**. This seems a good point to mention a further striking feature of some logical theories of information: a distinction between *positive* and *negative* information, and matching this, a simultaneous and mutually recursive use of 'support' and 'rejection' of formulas at pointed models in modal semantics. Examples of this approach are the 'data semantics' of Veltman (1984), the 'polarity-based' situation semantics of Barwise and Perry (1983), various systems in relevant logic (Omori & Wansing, 2019), and of course, there is also the three- (or higher-) valued tradition in intuitionistic and modal logic (Wansing, 1993). Similar ideas keep returning under new names in logical semantics, especially in the hyper-intensional logics discussed in Sect. 5.5.

There is something appealing to the positive/negative distinction, even in the restricted setting of mathematical inquiry. The modal-style negation $\neg\varphi$ in an intuitionistic model tells us that the current space of inquiry will never lead to support for $\varphi$: as in some indirect cultures, negation is signaled by silence. But one can also have a concrete counter-example to $\varphi$ available right now, representing a stronger form of negation as '*certified* exclusion'.

With a partial approach, logical operators split into variants, and so do notions of consequence, in ways that are well-known from the literature (Blamey, 2002). Such a framework also allows us to separate positive support from negative support in new ways. An example is the truthmaker semantics of Fine (2014) and later publications. It uses complete distributive lattices as semantic structures, and then interprets a propositional language with clauses like the following. An information state $s$ supports a disjunction $\varphi \vee \psi$ if $s$ is the infimum of states $t$ supporting $\varphi$ and $u$ supporting $\psi$. This is the attractive idea of a disjunction as a 'mixture' that has a distinguished history. Next, an information state $s$ rejects $\varphi \vee \psi$ if it rejects both $\varphi$ and $\psi$: which is the classical view of refuting both disjuncts, without adding rejection of mixtures.

**Translations**. Counteracting the diversity of richer partial logics over preorders or lattices is the existence of connections under translation. For instance, faithful translations from many-valued partial logics into classical true/false logics have long been folklore. On this basis, van Benthem (2018) presents a translation from Fine's truthmaker semantics into the above modal information logic with a double recursion

[23] I will not say much about Domain Theory in this paper, but it is rather surprising to me how little impact this fundamental and elegant theory of information and computing developed in the 1970s has had in philosophical logic or even in modal logic.

trick that goes back to Gilmore in the 1950s.[24] There might also be a Fernandez-style converse translation here, and many further translations no doubt exist in this realm. Even so, it would also be of interest to show negative results for a change, such as the non-translatability of the above two extended modal logics of information, the one for suprema/infima and that for update conditionals.

**Conclusion**. Preorder models and lattices support richer languages than those of intuitionistic logic or modal S4, and provide a vehicle for modal analyses of information combination or update. While this opens up an area of investigation, it is surprising how easy it is to define new simple modal systems for which open problems arise at once. The unknown starts right at our doorstep.

## 5.4 Possibility Semantics

**Motivations for partializing**. One may view the preceding sections as modalizing an approximation or growth relation $\leq$ in models for a propositional language. Let us now enrich the vocabulary, and start from a classical modal language with an arbitrary modality whose accessibility relation between points can stand for anything from epistemic uncertainty to time steps or process actions. Can we also, conversely, partialize this standard semantics to a setting that analyzes total worlds or outcomes as arising from ordered partial stages, without automatically committing to a shift to a non-classical base logic?[25]

Systems to this effect were proposed around 1980. The motivations in Humberstone (1981) for a partialized modal semantics were philosophical, such as approximating over-committed possible worlds in the classical sense by partial situations, or basing knowledge and belief on partial 'possibilities'. The motivations in van Benthem (1981) (published in van Benthem, 2016) for partializing the semantics for the language of first-order logic (with quantifiers viewed as modalities) were mathematical: removing the inelegant non-canonicity in the usual Henkin-type completeness argument which needs an arbitrary choice of a maximally consistent extension, and then removing or reducing the appeals to non-constructive principles like the Axiom of Choice in standard model theory. These possibility semantics for classical systems did not catch on widely in the ensuing decades, though rediscoveries have occurred, cf. Rumfitt (2015).

**Possibility semantics for modal logic**. Consider the language of propositional modal logic, with a universal modality $\Box$ and its existential dual $\Diamond$ whose standard meanings we wish to adjust to a stage-based setting. Modal models are now

[24] As observed earlier, translations need not reduce systems in a deeper sense, and key features of truthmaker logics, such as their tolerance of associative 'sup-completions' (Fine, 2014) need not follow from the connection with modal information logic.

[25] *Note*. The partiality in this section is not the many-valued notion discussed earlier.

tuples $\mathbf{M} = (W, R, \leq, V)$ where $\leq$ is a preorder and $R$ is a binary relation subject to the following two constraints, for all points $s, t, u$:

(a) if $s \leq t\ R\ u$, then $\exists v :\ s\ R\ v \leq u$,

(b) if $s\ R\ t$, then $\exists u \geq s\ \forall v \geq u\ \exists w \geq t\ :\ v\ R\ w$.

We also stipulate that the sets $V(p)$ assigned by the valuation to proposition letters are upward-closed and regular w.r.t. $\leq$ (cf. Sect. 5.2 for this notion).

Now the key truth conditions run as follows:

$\mathbf{M}, s \models \varphi \wedge \psi$ iff $\mathbf{M}, s \models \varphi$ and $\mathbf{M}, s \models \psi$

$\mathbf{M}, s \models \varphi \vee \psi$ iff for all $t \geq s$, there is a $u \geq t$ with $\mathbf{M}, u \models \varphi$ or $\mathbf{M}, u \models \psi$

$\mathbf{M}, s \models \neg\varphi$ iff for no $t \geq s$, $\mathbf{M}, t \models \varphi$

$\mathbf{M}, s \models \Box\varphi$ iff for all $t$ with $sRt$, $\mathbf{M}, t \models \varphi$[26]

The above two perhaps somewhat technical-looking conditions on possibility models belong to a family of possible connections between the accessibility relation $R$ and the partial extension relation $\leq$ in the literature. They are there to ensure that the following assertion becomes true: The denotations of any formula of the language are upward-closed and regular.[27]

The minimal modal logic K is the complete theory of possibility semantics in this sense. First, each standard modal model is a possibility model if we take the identity relation for $\leq$, and in that special case, standard modal semantics coincides with possibility semantics. Thus, non-theorems of K are refutable on possibility models. But also, given any pointed possibility model for a modal formula $\varphi$, we can turn this model into a classical modal model for $\varphi$ by restricting to the submodel of points that 'decide' all subformulas of $\varphi$.[28]

**Concerns and considerations**. Now all this invites discussion. The presented semantics looks somewhat engineered, so have we just managed to replace something simple (standard modal semantics) by something more complex (possibility semantics) for no better reason than ideological purity? One mathematical benefit of this semantic move is that the current more complex set-up allows for completeness proofs that are much more constructive than the usual ones, involving no appeals to set-theoretic equivalents of the Axiom of Choice like Zorn's Lemma. A philosophical benefit is that one can now model functional intuitions of modal notions that would not work in standard semantics, such as an agent's belief referring to just one possibility instead of a set of worlds.

---

[26] The existential modality is usually interpreted in possibility semantics by spelling out its classical definition $\neg\Box\neg\varphi$ with negation read in the new style.

[27] The intuitionistic modal logic of Bozic and Došen (1984) uses only the first condition.

[28] More precisely, for each relevant subformula $\alpha$, these special points make either $\alpha$ or its negation true. (This is stronger than the validity of $\alpha \vee \neg\alpha$ with disjunction read in the above sense.) For details, cf. the Filtration Lemma for possibility models in Holliday (2021). I thank Wes Holliday for removing some epicycles in my thinking about this completeness issue.

Another point to note is that the 'packaging' in this setting (more complex than in Kripke's semantics for intuitionistic logic) can be modified by shifting functions between the truth conditions, assumptions on denotations of formulas and structural constraints on the underlying possibility frames. In particular, no particular modal logic of possibility is favored by the possibilities setting, since there is a variety of natural meanings for the logical operations.[29]

**Modern themes in possibility logic**. A comprehensive development of the paradigm has been given by Holliday and co-workers (Holliday, 2021). This includes both philosophical themes, establishing connections to epistemology and philosophy of language, and mathematical ones, such as choice-free representation theorems in algebraic logic (Bezhanishvili & Holliday, 2020) or prospects for a choice-free model theory.[30] A general theme that returns here is *translation*. In the spirit of Gödel's translation of intuitionistic logic into S4, van Benthem et al. (2017) translates many modal possibility logics faithfully into bimodal logics over possibility frames, thus providing a richer modal realm where many different logics of possibility and accessibility coexist.

**Set-based models**. One theme in possibility semantics points ahead at Sect. 5.5 below, namely the existence of special models whose points are sets. A good method for developing intuitions about possibility models is through the following construction. One takes a standard relational model $\mathbf{M} = (W, R, V)$ and considers the powerset of $W$ with the inclusion relation and the natural lifting of the accessibility relation $R$ to sets:

$$XRY \text{ iff } \forall s \in X\, \forall t \text{ with } Rst : t \in Y.$$

Setting an appropriate valuation for proposition letters as sets of sets yields a model $poss(\mathbf{M})$ that satisfies the above conditions for a possibility model.[31]

Now a simple but important *Possibilization Lemma* due to Holliday states that for all propositional modal formulas $\varphi$,

$$poss(\mathbf{M}), X \models \varphi \text{ iff } \mathbf{M}, s \models \varphi \text{ for all } s \in X.$$

Thus, these models have what might be called a 'supervaluation' flavor. In particular, for each formula $\varphi$, there is a unique weakest possibility (largest set) that supports $\varphi$. General possibility models can then be seen as an abstraction out of these structures, where one must decide which properties of set-possibilities make sense in general, and which are just special technical features of sets.

---

[29] For instance, intuitionistic disjunction is immediate choice, classical disjunction is eventual choice, and the earlier-mentioned 'mixture' interpretation of disjunction makes sense too.

[30] This line continues the filter-based approach to ultraproducts and model-theoretic definability results based on these that was initiated in van Benthem (1981).

[31] Many semantics of this sort impose valuations of special forms guaranteeing closure of denotations for atomic propositions under subsets, or unions, or yet other properties.

Reversing this abstraction process, there is an issue of representing abstract possibility frames as concrete set models. This is not a simple topic, and the best available results are only partial, cf. Harrison-Trainor (2017).

**Temporal dynamics**. Possibility models can contain information states about more than one world or object: states can be incompatible in many ways. A concrete form of modal possibility semantics occurs with *dynamical systems* where the relation $R$ is a transition function $F$ that takes states of the system to other states. Here the state space carries a topology whose open sets represent information pieces or possible outcomes of available experiments, in the style of Vickers (1989), Kelly (1996). This richer setting adds new issues. For instance, from an informational viewpoint, *continuous* functions are now important, as these approximate function values through approximations of their arguments, thus providing a link with computation. In the special case of pre-order topologies, this just means that the function is *monotonic* for inclusion:

if $x \leq y$, then $F(x) \leq F(y)$.

These functions provide an abstract notion of *computional process*, an obvious dynamic addition to the logical study of static information structures.

Just as an illustration, the following axiom for continuity occurs widely in the modal literature on topological systems (Kremer & Mints, 2007):

$$\langle F\rangle[\leq]\varphi \rightarrow [\leq]\langle F\rangle\varphi.$$

This commutation principle is not valid in the above general modal possibility logic, but it singles out an important application area.[32]

**Conclusion**. The informational view of modal models combines naturally with accessibility relations that can stand for a wide range of notions, including dynamical systems and computation using information structure for some purpose. It is somewhat surprising that modal-style analysis of these natural dynamic companions to widely studied static notions has taken so long.

## 5.5 Set Lifting

We noted in the preceding section that one concrete way of creating partial possibilities models is by set lifting standard semantics. Set lifting keeps emerging with a certain frequency in the logical literature as a means of providing richer partialized semantic structures, but it does not seem to have a well-known established tradition. In this section, we survey a few manifestations.

[32] A similar modal perspective applies to Domain Theory, but there, two notions of continuity occur, our monotonicity version and 'Scott continuity' (Abramsky & Jung, 1994).

**Temporal Interval Logic**. Around 1980, a number of authors proposed changing the usual Prior-style semantics of temporal logics in terms of durationless points to temporal interval structures where logical languages could be interpreted directly at suitable sets of points. Kamp (1979) presented a discourse semantics for temporal expressions in natural language where discourse creates event structures, where events have intervals as their running times. van Benthem (1979) proposed both abstract and set-lifted interval semantics for temporal logic as representing ontologically better motivated primitives, more in the tradition of mereology. Interestingly, as an independent discovery around the same time, intervals were also proposed in the AI literature on 'common sense physics' (Allen, 1983; Hayes, 1979), as providing a model for databases for maintaining temporal information that admits of simpler computation than models closer to standard mathematical point structures.

Here is a sample of a set-based interval tense logic, cf. Cresswell (1977), Halpern and Shoham (1986), Venema (1991) for systems with further vocabulary. An *interval model* $\mathbf{M} = (T, <, \mathbf{I}, V)$ consists of an ordered set $(T, <)$ with a family $\mathbf{I}$ of subsets of $T$, and a valuation $V$ for proposition letters. There may be further intuitive constraints, say, making intervals *convex sets* $X$ satisfying $\forall s, t \in X : \forall u : s < u < t \Rightarrow u \in X$. The following truth definition lifts the usual semantics of temporal logic to intervals as indices of evaluation, while adding a modality for the new relation of inclusion. We only display the major clauses, which reflect intuitions about temporal statements in natural language when the primitive components of time are intervals with duration:

$$
\begin{array}{ll}
\mathbf{M}, X \models \varphi \wedge \psi & \text{iff } \mathbf{M}, X \models \varphi \text{ and } \mathbf{M}, X \models \psi \\
\mathbf{M}, X \models \neg\varphi & \text{iff for no } Y \subseteq X, \mathbf{M}, Y \models \varphi \\
\mathbf{M}, X \models \varphi \vee \psi & \text{iff there exist } Y, X \text{ with } X = Y \cup Z \\
 & \text{such that } \mathbf{M}, Y \models \varphi \text{ and } \mathbf{M}, Z \models \psi \\
\mathbf{M}, s \models F\varphi & \text{iff for some } Y >> X, \mathbf{M}, Y \models \varphi \\
 & \text{with } Y >> X \text{ for } \forall x \in X,\ y \in Y : x < y \\
\mathbf{M}, X \models [\downarrow]\varphi & \text{iff for all } Y \subseteq X, \mathbf{M}, Y \models \varphi
\end{array}
$$

Typically, standard logical notions split in the more structured interval setting. For instance, the above interval disjunction is the merge disjunction of modal information logic, but the language can also define the cofinal choice disjunction of possibility semantics (reading $Y \subseteq X$ as a counterpart of the earlier $x \leq y$). But there are also differences with possibility semantics, since the temporal setting brings its own features, such as the special set lifting $X << Y$ for the natural intuitive notion of total precedence of intervals.

Another difference are new constraints on intervals that have no counterpart for durationless temporal points, or for possibilities as usually conceived. One intriguing candidate is 'Homogeneity' (van Benthem, 1983): all intervals are isomorphic, an ontological view of time where the large reflects in the small.

Also differently from earlier logics, the denotations of propositions are not required to have the special properties of upward-closure or regularity. Upward clo-

sure is downward closure in the interval subset order, and this only holds for special temporal assertions such as "being alive". Instead, interval logic offers a general theory which includes reasoning about propositions with many different kinds of temporal behavior coming, e.g., from well-known linguistic classifications of verb classes into states, activities, or accomplishments.

**Other set liftings**. Set lifting is also common in other areas, especially computer science, witness the theory of power domains (Plotkin, 1976), concurrent dynamic logic (Peleg, 1985), or event structures (Winskel, 1989). Brink (1993) presents a mathematical theory of set lifting motivated by a range of applications all the way from computation to verisimilitude in the philosophy of science.

In a more logical setting, a modern case of set lifting is the logic of *dependence* of Väänänen (2007), going back to Hodges (1997), where the language of first-order logic is interpreted on sets of assignments in order to guarantee a compositional semantics for Hintikka's IF logic. This lifts the propositional connectives in the same way as we saw for temporal interval logic, while also retaining earlier persistence constraints. But crucially, the new richer structure is also exploited in other ways, adding new forms of quantification as well as dependence atoms making the functional dependence structure in sets of assignments explicit.[33]

A final, quite recent case of set lifting is the 'bilateral semantics' of Aloni (2019) for classical modal logic which redraws the boundaries of semantics and pragmatics in order to deal with delicate linguistic phenomena such as free choice permissions. Unlike possibility semantics, this semantics uses support and rejection clauses on sets of classical worlds, lifting the accessibility relation in new ways. Beyond persistence, another constraint on propositions is that the sets supporting (or rejecting) a formula are closed under unions.[34]

Thus, motivations for set lifting, old and new, are diverse: what is common to many cases is the richer pattern of interpretation for basic logical operations.

**Theory of complex algebras**. There exists an unfortunately little-known mathematical theory of set lifting. Given any algebra $\mathbf{A}$, define the *complex algebra* $C(\mathbf{A})$ as the power set of the set of objects in $\mathbf{A}$, lifting the operations $f$ as follows, where we just display the case of a binary operation for simplicity:

$$X\, C(f)\, Y = \{xfy \mid x \in X, y \in Y\}$$

[33] Dependence logic has various connections with inquisitive logic (Ciardelli et al., 2018), one more framework with many points of contacts with the themes in this paper. For the special case of set-lifting, especially relevant is a precursor from the field of intermediate logics in between classical and intermediate logics, namely the 'logic of finite problems' of Medvedev (1966), whose decidability is one more open landmark challenge.

[34] One of Aloni's systems does not just set lift, but combines interpretation on worlds and on sets of worlds. This joint way of proceeding seems closer to the two-sorted approach in terms of 'objects' and 'types' in Barwise and Seligman (1995), and it might also suggest a generalization of modal neighborhood semantics (Pacuit, 2017) to combined point- and set-languages.

An early issue investigated at this level of generality is which equations valid for the original algebras remain valid as they stand for their lifted versions. A theorem in Gautam (1957) says this is rare, being the case only if the terms in the equations have only single occurrences of variables. Even so, other styles of axiomatization remain possible in principle. Further theory of complex algebras from a modal perspective can be found in Goldblatt (1989), Brink (1993). As for more recent developments, Hodkinson et al. (2001) is a modern source using game techniques to axiomatize varieties of complex algebras in a perspicuous manner when the original algebras have a computably enumerable theory. Incidentally, the theory of complex algebras covers both full power set algebras and subalgebras generated by suitably closed subfamilies of sets.

To understand what complex algebras do, and how they relate to the topics discussed earlier, it helps to think of a concrete example: complex Boolean algebras. Moving to a power set, or a suitable subfamily thereof, we find lifted operations that one might call the *inner Booleans* of the complex algebra:

$$X + Y = \{x + y \mid x \in X, y \in Y\}$$
$$X.Y = \{x.y \mid x \in X, y \in Y\}$$
$$-X = \{-x \mid x \in X\}.$$

Note that inner Boolean disjunction is precisely the set-lifted disjunction of temporal interval logics or dependence logics. In addition to inner Booleans, there are also *outer Booleans*, as one can also interpret disjunction or negation in the usual way via their corresponding set-theoretic operations. This duplication of available meanings for logical operations extends to the basic constants. For instance, the zero of Boolean algebra splits into an inner Boolean $\{\emptyset\}$ and an outer Boolean $\emptyset$ standing for the empty family of sets.[35]

Viewed more abstractly, the binary modality $\langle \text{inf} \rangle$ of Sect. 5.3 is a generalized counterpart to inner Boolean disjunction, and it also illustrates concretely how one can think of power algebras as supporting *modal operators*.

An intriguing feature of complex algebra that might extend to its modern manifestations are natural, yet highly non-trivial open problems. For instance, no simple purely modal axiomatization is known for complex Boolean Algebra as introduced above, though (Goranko & Vakarelov, 1999) found a workable axiomatization by adding special rules for an additional 'difference modality' that can then be translated back into the pure language. Moreover, easy to understand, and yet vexing upon reflection, is the complex algebra counterpart of a problem posed for modal information logic in Sect. 5.3: *Is validity in complex Boolean Algebra decidable?* This has been unresolved for several decades now, and it might pose a real challenge to existing methods for establishing decidability.

**Set lifting and abstract semantics**. While set lifting is an elegant way of creating richer partialized versions of given algebras, one can also move on to more abstract versions. For instance, in temporal interval logic, an alternative approach

[35] This splitting of constants is common in hyperintensional semantics, cf. Leitgeb (2019).

treats intervals as primitive objects ('chunks of time') with two relations: one for temporal precedence, and one reflecting the extended nature of intervals: inclusion as in the above, overlap (Röper, 1980; Wiener, 1914), or yet others. This leads to issues of representation of abstract interval or event models by concrete set structures that go back to Russell's views on the interface of science and common sense (Russell, 1926). van Benthem (1983) has an extensive representation theory connecting the two views of time both ways.[36] Likewise, in possibility semantics, we noted the issue of set construction versus abstract structures that adopt some, but not necessarily all properties that hold for set-theoretic reasons. Similar abstract structures are emerging for the more recent manifestations of set lifting that we have mentioned, as people seek greater generality, but discussing these would take us too far here.[37]

**Conclusion**. Set-lifting is a general way of creating richer models for classical systems that keeps returning in linguistic and philosophical semantics, and its mathematical theory offers a systematic explanation for many features in the current literature. The connection between set lifting, which imports specific structure of power sets, and more general abstract structures is intriguing, and there is room for a general representation theory behind its many manifestations in the current literature. This may well call for a new phase of mathematization in modal logic going beyond the first wave of the 1970s, or at least, for deepening the contacts between the scattered world of applied modal logic today and fundamental developments in algebraic logic and co-algebra.

## 5.6 Conclusion

The introduction to this paper stated my commitment to relational structure as the locus of modality. How one views the semantics is of course just one aspect of modal logic, and there are further mantras of the Amsterdam School, concerning modal languages and complexity, that I have refrained from reciting. I also stressed abstract realms as a good place to work, facilitating pleasant and unforced walks from one logical system to another, borrowing ideas across.

The case history presented here of partialized or set-lifted relational models may have shown that modal language and logic design is alive and well, and in particular that, when the dust has settled, current hyperintensional semantics is not a threat to, but a part of modal logic. More generally, in my mathematical perspective, seemingly competing frameworks in philosophical logic show great similarities, while in a

[36] An interesting analysis of Russell's views of 'public time' as a limiting point structure arising out of growing interval-based 'private times' is given in Thomason (1989) using an inverse limit construction that might make sense for all the partial semantics we have considered.

[37] Within the confines of this already quite packed survey, it has not been possible to include an account of *coalgebra* (Jacobs, 2016) as an abstract approach to set lifting. We confine ourselves to mentioning the co-algebraic technique of *predicate lifting* (Moss, 1999; Pattinson, 2003) as a general unifying approach that ties together many existing modal logics.

historical perspective, their communities are engaged in shared endeavors. Finally, in these pursuits, modal logic and philosophical logic generally can only profit from an openness to insights from fields as diverse as philosophy, linguistics, mathematics, and computer science.

At the start of this paper, I objected to questions like "What are possible worlds?" as being intellectually misguided. Nevertheless, despite having used the word "modal" many times in this text, I admit there is a reasonable question of "What makes topics modal?" Even so, I have not offered an answer, let alone a definition. Overall, I feel that broad history of ideas is more revealing of a field than a priori definition and boundary setting.[38] And there is not even one such history. If you wish, modal logic is a Sea of Stories (Rushdie, 1990).

Which brings me to a final question from my students which I do find to the point. "Your lecture was nice, but: does it *always* have to be modal logic?" To that question, I do have a definite answer: "Not necessarily".

**Acknowledgements** I thank Wes Holliday for thoughtful comments and firm corrections to a first draft of this paper. I thank the two referees for this volume for painstaking reports including the opening of some windows to further relevant perspectives. Finally, the audience at the Saul Kripke Seminar in New York (November 2021) provided helpful feedback at a late stage, in particular, Kit Fine, Allen Hazen, Albert Visser, Heinrich Wansing and Yale Weiss.

## Appendix

## Axiomatizing Modal Information Logic

While axiomatic proof systems for the sparse language of K or S4 are well-known, working with the richer modal languages introduced in this paper may be less familiar. This appendix sketches a Henkin-style heuristics for finding an axiomatization of the modal information logic of Sect. 5.3. The language uses two mild additional devices: *nominals* naming individual points, and a global *existential modality* $E\varphi$ saying that $\varphi$ is true in at least one point in the current model. plus its universal dual. We only treat the case of the binary modality $\langle \mathrm{inf} \rangle \varphi\, \psi$, written here for convenience in infix notation as $\varphi + \psi$. Key parts of the axiomatization will be found in the course of a step-by-step heuristic canonical model-style analysis, for details see Blackburn et al. (2001).

(a) Standard background: (i) the axioms and proof rules of the minimal modal logic for a binary modality $\varphi + \psi$ plus all modalities defined in what follows, (ii) the axioms of S5 for the global modality, as well as its usual connections to all other modalities, (iii) all standard proof principles from hybrid logic for nominals and the

[38] van Benthem (2006) offers a history of ideas approach to the role of logic in philosophy.

existential modality, (iv) two obvious principles for $\langle \text{inf} \rangle$: commutativity $(\varphi + \psi) \rightarrow (\psi + \varphi)$ and generalized reflexivity $\varphi \rightarrow (\varphi + \varphi)$.

We will add further axioms and rules as needed.

(b) Define a unary modality $\langle \uparrow \rangle \varphi$ as $\varphi + \top$. Its matching abstract order $\leq$ in a canonical Henkin model is reflexive, since the principle $\varphi \rightarrow (\varphi + \top)$ is derivable from the above generalized reflexivity axiom. Next, to make the order transitive, we add weak associativity:

$((\varphi + \top) + \top) \rightarrow (\varphi + \top)$[39]

(c) In the canonical model of all maximally consistent sets, we want to (c1) name worlds, (c2) introduce witnesses for the two arguments of the abstract existential binary modality $\varphi + \psi$, and (c3) make the abstract ternary relation $Rzxy$ for $\langle \text{inf} \rangle$ in a standard model completeness proof coincide with the real infimum relation $inf(x, y) = z$ with respect to the binary order $\leq$ in canonical models induced by the modality $\langle \uparrow \rangle$ in the usual way.

(c1) To name worlds, we make a standard use of the hybrid proof rule

If $\vdash p \rightarrow \varphi$, where the nominal $p$ does not occur in $\varphi$, then $\vdash \varphi$

This allows adding a fresh nominal to max-cons sets under construction.

(c2) In a similar manner, to create witnesses for the components of our binary modality, we employ one more familiar hybrid proof rule:

If $\vdash (E(k \wedge \varphi) \wedge E(m \wedge \psi) \wedge (k + m)) \rightarrow \alpha$, where the nominals $k, m$ do not occur in $\varphi, \psi, \alpha$, then $\vdash (\varphi + \psi) \rightarrow \alpha$

(c3) will be dealt with presently in step (e).

(d) Define the standard modal relation $\Sigma \leq \Delta$ on maximally consistent sets just in case $\alpha \in \Delta$ implies $\langle \uparrow \rangle \alpha \in \Sigma$ for all formulas $\alpha$.

(e) We show that, for any three max-cons sets, if $k + m \in \Sigma$, $k \in \Delta$, $m \in \Gamma$, then $\Sigma$ is a greatest lower bound of $\Delta$ and $\Gamma$ in the just-defined ordering $\leq$. By our construction of the canonical model, we can assume that $\Sigma$ is named by some nominal $n$. This is the step where we need enough axioms to reflect the definition of infima.

(e1) First, we show the lower bound property, starting with the case $\Sigma \leq \Delta$. Consider any formula $\alpha \in \Delta$: we then have $E(k \wedge \alpha)$ present in any max-cons set. At this point, to proceed, it suffices to have

$((k + m) \wedge E(k \wedge \alpha)) \rightarrow (\alpha + \top)$

The consequent, which equals $\langle \uparrow \rangle \alpha$, is then in the set $\Sigma$, as required. But this formula can be derived in our system: $E(k \wedge \alpha)$ implies $U(k \rightarrow \alpha)$ (with $U$ the global universal modality) in hybrid logic, and together with $k + m$, this implies $\alpha + m$ and

[39] Weak associativity holds on all pre-order models. Full associativity requires an existence condition for infima which can induce undecidability (Kurucz et al., 1995).

hence also $\alpha + \top$. The same argument works to show that $\Sigma \leq \Gamma$, now also using the symmetry axiom for the + modality.

(e2) Next, we show that $\Sigma$ is a greatest lower bound. Let $\Sigma$' be a lower bound of $\Delta$ (named by the nominal $k$) and $\Gamma$ (named by $m$) in the order $\leq$. We must show that $\Sigma' \leq \Sigma$. By the definition of $\leq$, we have $\langle\uparrow\rangle k, \langle\uparrow\rangle m \in \Sigma'$. Here we need a further axiom to proceed. The following holds in all pre-orders:

$$(\langle\uparrow\rangle k \wedge \langle\uparrow\rangle m \wedge E(n \wedge (k + m))) \rightarrow \langle\uparrow\rangle n$$

and we simply incorporate it into our logic [this is the 'heuristic' mentioned above]. We conclude that $\langle\uparrow\rangle n \in \Sigma'$. From this, it follows that, for any formula $\alpha$, if $\alpha \in \Sigma$, then $E(n \wedge \alpha) \in \Sigma'$ and reasoning as before, $\langle\uparrow\rangle\alpha \in \Sigma'$.

This concludes our analysis of modal information logic on pre-orders.[40]

As for stronger proof systems, one can build more into this analysis. Using nominals, one extra axiom makes $\leq$ a partial order. One can also impose full associativity for +, or add Aloni's union-closure condition $(\varphi + \varphi) \rightarrow \varphi$ to turn the Henkin model into a special kind of pre-order. The analysis also extends to possibility logic with standard modalities $\Box$, $\Diamond$ with their own accessibility relation $R$ by supplying axioms that match natural constraints on $R$, $\leq$ such as $\langle\uparrow\rangle\Diamond\varphi \rightarrow \Diamond\varphi$ for $x \leq y\, R\, z \Rightarrow x\, R\, z$ and $\Diamond\langle\uparrow\rangle\varphi \rightarrow \Diamond\varphi$ for $x\, R\, y \leq z \Rightarrow x\, R\, z$.

Is there also a purely modal complete proof system without hybrid gadgets? This raises interesting issues in the complex algebra introduced in Sect. 5.5.

## Modal and Metamodal Languages

As stated at the beginning, we have not emphasized the role of modal syntax per se in this paper, although that, too, is an important part of the appeal of modal analysis of reasoning practices. Here are just a few thoughts connecting syntax issues to our relationalist stance.

**Modal languages and focus on single worlds**. Modal languages are usually 'local', describing the structure of a model from the vantage point of one current point. This may be one source of the idea that possible worlds encode all the information we need about relational models. The local design allows for a simple analysis of completeness with canonical models where points are maximally consistent sets of formulas, while a simple choice of accessibility relation makes sure that the different local descriptions match up correctly. In canonical models, unlike in the world of professional soccer, a collection of overpaid self-centered stars makes a good team. What does this mean?

[40] The above proof leaves out some easy details for dealing with the existential modality.

Modal formulas $\varphi$ describe properties of points in models up to worlds reachable at a distance matching the modal depth of $\varphi$. To describe models in full depth, we need an *infinitary modal language* with infinite conjunctions and disjunctions of sets of formulas. For each pointed model $\mathbf{M}, s$ there is an infinitary modal formula (a 'modal Scott sentence') which holds in exactly those models $\mathbf{N}, t$ that have a bisimulation with $\mathbf{M}, s$. The construction of modal Scott sentences describes, inductively through the ordinals, using a suitable $\Diamond, \Box$ pattern, which types of points are reachable at preceding ordinal description levels.[41]

In a sense, then, reading off the modal properties of a single point tells us the structure of a whole model, insofar as reachable from that point. But these descriptions would not be true without the relational pattern, so this seems no more surprising than that pools surrounded by trees reflect those trees.

Even so, the canonical model construction for modal logic is remarkable in that it produces one *universal model* for all consistent sets of formulas. This construction by '1-types' has no counterpart for first-order logic, unless one moves to the possibility semantics of van Benthem (2016), Holliday (2021).

A related way in which the syntax of the modal language is special shows in the method of *filtration* for proving decidability. Given a finite set of formulas and a relational model, we can extract the types of worlds with respect to just that finite set, and turn them into a model respecting the truth values of the selected formulas. While filtration can be defined for first-order logic (van Benthem & Bezhanishvili, 2021), it seldom yields models for the relevant formulas, and it is even undecidable when this happy coincidence occurs.

**Modal logics for meta-purposes**. The two higher levels of relationalism mentioned in the Introduction, too, fall within the realm of modal logic: provided we are willing to extend the syntax once more.

First, we suggested that abstract patterns are best understood by looking at models in their relations to other models. As it happens, there is a growing variety in the recent literature of logics with cross-model modalities, interpreted in the following pattern:

> $\mathbf{M}, s \models [\bullet]\varphi$ iff $\mathbf{N}, s \models \varphi$ for all models $\mathbf{N}$ (perhaps from some relevant family of models) standing in some relation $R^\bullet$ to $\mathbf{M}$

Relevant relations $R^\bullet$ include those of being a submodel, homomorphic image, bisimulation image, filtration, and the like. Barwise and van Benthem (1999) used this setting for a study of 'entailment along a relation' as a generalized form of consequence involving information transfer between models. Broad families of such logics include dynamic-epistemic logics, where shifts between models correspond to information updates (Baltag & Renne 2016), logics for graph changes in computation or games, and logics that formalize part of the model theory of standard modal logic (van Benthem & Bezhanishvili 2021). Such logics for model change can be studied with standard modal techniques.

[41] For finite models, Scott sentences can even be defined in propositional dynamic logic.

A second suggestion in our Introduction was that translations between languages and logics provide another relational structure that can itself be profitably modalized. Such modalizations have happened in the specialized setting of arithmetical proof systems, witness the 'interpretability logic' surveyed in Japaridze and deJongh (1998), many of whose principles also make sense for general translations and relative interpretations between arbitrary theories. Still more generally, one can also introduce modalities inside one language and proof systems that refer to principles statable and provable in another language via some definable translation. This sort of natural combination of different systems is central to the 'combining logics' framework of Gabbay (1999).

*Added in Print*. In the time since this paper started circulating, several of the stated open problems have been settled by Søren Brinck Knudstorp. Decidability and axiomatization for modal information logic including its natural implication modality are proved in Brinck Knudstorp (2023b). Connections with several truthmaker logics have been established in Brinck Knudstorp (2023a). As for set-lifted logics, a still unpublished result is the following. Given any semilattice $(\mathbf{S}, o)$, consider its complex algebra for the operation $o$ with the usual Boolean set operations added. Then for any class $\mathbf{S}$ of semilattices that contains the semilattice $(P(N), \cup)$, where $N$ is the natural numbers, the equational theory of the variety generated by its complex algebras is undecidable. Two consequences are that modal information logic over the special class of semilattices is undecidable, and that set-lifted Boolean Algebra is undecidable. By a refinement of the techniques involved, it can also be shown that Urquhart's relevant logic $\mathbf{S}$ is undecidable, a result to be presented at LICS 2024.

## References

Abramsky, S., & Jagadeesan, R. (1994). Games and full completeness for multiplicative linear logic. *Journal of Symbolic Logic,59*(2), 543–574.

Abramsky, S., & Jung, A. (1994). Domain theory. In S. Abramsky, D. Gabbay, & T. Maibaum (Eds.), *Handbook of logic in computer science* (Vol. III, pp. 1–168). Oxford University Press.

Allen, J. (1983). Maintaining knowledge about temporal intervals. *Communications of the ACM, 26*, 832–843.

Aloni, M. (2019). *Pragmatic enrichments in bilateral state-based modal logic*. Paper presented at the Truthmaker Semantics Workshop: ILLC, University of Amsterdam.

Anderson, A., Belnap, N., & Dunn, J. (2017). *Entailment, the logic of relevance and necessity* (Vol. 2). Princeton University Press.

Aucher, G. (2014). Dynamic epistemic logic as a substructural logic. In A. Baltag & S. Smets (Eds.), *Johan van Benthem on logic and information dynamics* (pp. 855–880). Springer.

Baltag, A., & Renne, B. (2016). Dynamic-epistemic logic. *Stanford Electronic Encyclopedia of Philosophy*.

Barrett, T., & Halvorson, H. (2016). Morita equivalence. *The Review of Symbolic Logic, 9*, 556–582.

Barwise, J. (1989). *The situation in logic*. CSLI Publications.

Barwise, J., & Perry, J. (1983). *Situations and attitudes*. The MIT Press.

Barwise, J., & Seligman, J. (1995). *Information flow. the logic of distributed systems*. Cambridge University Press.

Barwise, J., & van Benthem, J. (1999). Interpolation, preservation, and pebble games. *Journal of Symbolic Logic, 64*, 881–903.
Beth, E. (1956). Semantic construction of intuitionistic logic. *Mededelingen Koninklijke Nederlandse Akademie van Wetenschappen, 19*(11), 357–388.
Beth, E. (1959). *The foundations of mathematics. A study in the philosophy of science*. North-Holland.
Bezhanishvili, G., & Holliday, W. (2019). A semantic hierarchy for intuitionistic logic. *Indagationes Mathematicae, 30*, 403–469.
Bezhanishvili, N., & Holliday, W. (2020). Choice-free Stone duality. *Journal of Symbolic Logic, 85*, 109–148.
Bimbó, K., & Dunn, M. (2008). *Generalized Galois logics. Relational semantics of nonclassical logical calculi*. CSLI Publications.
Blackburn, P., de Rijke, M., & Venema, Y. (2001). *Modal logic*. Cambridge University Press.
Blamey, S. (2002). Partial logic. In D. Gabbay & F. Guenthner (Eds.), *Handbook of philosophical logic* (pp. 261–353). Springer.
Božic, M., & Došen, K. (1984). Models for normal intuitionistic modal logics. *Studia Logica, 43*, 217–245.
Brinck Knudstorp, S. (2023a). Logics of truthmaker semantics: Comparison, compactness, and decidability. *Synthese, 202*(6), 1–18.
Brinck Knudstorp, S. (2023b). Modal information logics: Axiomatizations and decidability. *Journal of Philosophical Logic, 52*, 1723–1766.
Brink, C. (1993). Power structures. *Algebra Universalis, 30*, 177–216.
Ciardelli, I., Groenendijk, J., & Roelofsen, R. (2018). *Inquisitive semantics*. Oxford University Press.
Cresswell, M. (1977). Interval semantics and logical words. In C. Rohrer (Ed.), *On the logical analysis of tense and aspect* (pp. 7–29). Gunter Narr Verlag.
Fernández Duque, D. (2006). A polynomial translation of S4 into intuitionistic logic. *Journal of Symbolic Logic, 71*(3), 989–1001.
Fine, K. (2014). Truth-maker semantics for intuitionistic logic. *Journal of Philosophical Logic, 43*, 549–577.
Fitting, M. (2007). Modal proof theory. In P. Blackburn, J. van Benthem, & F. Wolter (Eds.), *Handbook of modal logic* (pp. 85–138). Elsevier.
French, R. (2011). *Translational embeddings in modal logic* (Unpublished doctoral dissertation). Department of Philosophy, Monash University.
Friedman, H., & Flagg, R. (1986). Epistemic and intuitionistic formal systems. *Annals of Pure and Applied Logic, 32*, 53–60.
Gabbay, D. (1999). *Fibring logics*. Oxford University Press.
Gärdenfors, P., & Rott, H. (1995). Belief revision. In D. Gabbay, C. Hogger, & J. Robinson (Eds.), *Handbook of logic in AI and logic programming* (pp. 35–132). Oxford University Press.
Gautam, N. (1957). The validity of equations of complex algebras. *Archive für Mathematische Logik und Grundlagen der Mathematik, 3*, 117–124.
Gödel, K. (1933). Eine interpretation des intuitionistischen aussagenkalküls. *Ergebnisse eines Mathematischen Kolloquiums, 4*, 39–40.
Goldblatt, R. (1989). Varieties of complex algebras. *Annals of Pure and Applied Logic, 44*, 173–242.
Goranko, V., & Otto, M. (2006). Model theory of modal logic. In J. van Benthem, P. Blackburn, & F. Wolter (Eds.), *Handbook of modal logic* (pp. 249–329). Elsevier.
Goranko, V., & Vakarelov, D. (1999). Hyperboolean algebras and hyperboolean modal logic. *Journal of Applied Non-Classical Logics, 9*, 345–368.
Goré, R. (2000). Dual intuitionistic logic revisited. In *Proceedings Automated Reasoning with Analytic Tableaux and Related Methods* (pp. 252–267). Springer.
Goré, R., & Thomson, J. (2019). A correct polynomial translation of S4 into intuitionistic logic. *Journal of Symbolic Logic, 84*, 439–451.

Halpern, J., & Shoham, Y. (1986). A propositional modal logic of time intervals. In *Proceedings of the First Symposium on Logic in Computer Science* (lics) (pp. 279–292).
Harrison-Trainor, M. (2017). *A Representation Theorem for Possibility Models* (Unpublished doctoral dissertation). Group in Logic and Methodology of Science, University of California at Berkeley.
Hayes, P. (1979). The naive physics manifesto. In D. Michie (Ed.), *Expert systems in the microelectronic age*. Edinburgh University Press.
Hodges, W. (1997). Compositional semantics for a language of imperfect information. *Logic Journal of the IGPL, 5*(4), 539–563.
Hodkinson, I., Mikulás, S., & Venema, Y. (2001). Axiomatizing complex algebras by games. *Algebra Universalis, 46*, 455–478.
Holliday, W. (2021). Possibility semantics. In A. Daghighi (Ed.), *Research trends in contemporary logic*. College Publications.
Humberstone, L. (1981). From worlds to possibilities. *Journal of Philosophical Logic, 10*, 313–339.
Jacobs, B. (2016). *Introduction to coalgebra*. Cambridge University Press.
Japaridze, G., & de Jongh, D. (1998). The logic of provability. In S. Buss (Ed.), *Handbook of proof theory* (pp. 475–546). Elsevier.
Kamp, H. (1979). Events, instants and temporal reference. In R. Bäuerle, U. Egli, & A. von Stechow (Eds.), *Semantics from different points of view* (pp. 376–418). Springer.
Kelly, K. (1996). *The logic of reliable inquiry*. Oxford University Press.
Kocurek, A. (2018). *What can you say? measuring the expressive power of languages* (Unpublished doctoral dissertation). Group in Logic and Methodology of Science, University of California at Berkeley.
Kremer, P., & Mints, G. (2007). Dynamic topological logic. In M. Aiello, I. Pratt- Harman, & J. van Benthem (Eds.), *Handbook of spatial logics* (pp. 565–606). Springer.
Kripke, S. (1963). Semantical considerations on modal logic. *Acta Philosophica Fennica, 16*, 83–94.
Kripke, S. (1965). Semantical analysis of intuitionistic logic I. In M. Dummett & J. Crossley (Eds.), *Formal Systems and Recursive Functions: Proceedings of the Eighth Logic Colloquium* (pp. 92–130). North Holland.
Kurtonina, N., & Moortgat. (2009). Relational semantics for the Lambek-Grishin calculus. In *The mathematics of language, mol 2007* (pp. 210–222).
Kurucz, A., Németi, I., Sain, I., & Simon, A. (1995). Decidable and undecidable logics with a binary modality. *Journal of Logic, Language, and Information, 4*(3), 191–206.
Leitgeb, H. (2019). HYPE: A system of hyperintensional logic. *Journal of Philosophical Logic, 48*, 305–405.
Lewis, D. (1973). *Counterfactuals*. Blackwell.
Lewis, D. (1970). General semantics. *Synthese, 22*, 18–67.
Medvedev, Y. (1966). Interpretation of logical formulas by means of finite problems. *Doklady Akademii Nauk SSSR, 169*(1), 20–23.
Moisil, G. (1942). Logique modale. *Disquisitiones Mathematicae et Physicae, 2*, 3–98.
Moss, L. (1999). Coalgebraic logic. *Annals of Pure and Applied Logic, 96*(1–3), 277–317.
Negri, S. (2011). Proof theory for modal logic. *Philosophy Compass, 6*(8), 523–538.
O'Hearn, P., & Pym, D. (1999). The logic of bunched implications. *The Bulletin of Symbolic Logic, 5*(2), 215–244.
Olkhovikov, G. (2013). Model-theoretic characterization of intuitionistic propositional formulas. *Review of Symbolic Logic, 6*, 348–365.
Omori, H., & Wansing, H. (Eds.). (2019). *New essays on Belnap-Dunn logic* (Vol. 418). Springer.
Pacuit, E. (2017). *Neighborhood semantics for modal logic*. Springer.
Paterson, A. (1997). Bisimulation and propositional intuitionistic logic. In *Proceedings Concur 1997* (pp. 347–360).
Pattinson, D. (2003). Coalgebraic modal logic: Soundness, completeness and decidability of local consequence. *Theoretical Computer Science, 309*(1–3), 177–193.

Peleg, D. (1985). Concurrent dynamic logic. In *Stoc 1985: Proceedings of the Seventeenth Annual ACM Symposium on Theory of Computing* (pp. 232–239).
Plotkin, G. (1976). A powerdomain construction. *SIAM Journal of Computing, 5*, 452–486.
Rauszer, C. (1977). Applications of Kripke models to Heyting-Brouwer logic. *Studia Logica, 36*, 61–71.
Reynolds, J. (2002). Separation logic: A logic for shared mutable data structures. In *Lics 2002: Proceedings of the 17th Annual IEEE Symposium on Logic in Computer Science* (pp. 55–74).
Röper, P. (1980). Intervals and tenses. *Journal of Philosophical Logic,9*, 451–469.
Rumfitt, I. (2015). *The boundary stones of thought. An essay in the philosophy of logic*. Oxford University Press.
Rushdie, S. (1990). *Haroun and the sea of stories*. Granta Books.
Russell, B. (1926). *Our knowledge of the external world*. Allen and Unwin.
Stalnaker, R. (1968). A theory of conditionals. In N. Rescher (Ed.), *Studies in logical theory* (pp. 98–112). Blackwell.
Thomason, S. (1989). Free construction of time from events. *Journal of Philosophical Logic, 18*, 43–67.
Troelstra, A., & van Dalen, D. (1988). *Constructivism in mathematics* (Vol. 1). Elsevier.
Urquhart, A. (1972). Semantics for relevant logics. *Journal of Symbolic Logic, 37*(1), 159–169.
Urquhart, A. (2016). Relevance logic: Problems open and closed. *The Australasian Journal of Logic, 13*(1), 11–20.
Väänänen, J. (2007). *Dependence logic*. Cambridge University Press.
van Benthem, J. (1979). Points and periods. In *Proceedings Colloquium on Tense and Quantifiers*. University of Stuttgart.
van Benthem, J. (1981). *Possible worlds semantics for classical logic* (Technical Report No. ZW 8018). Mathematical Institute, University of Groningen.
van Benthem, J. (1983). *The logic of time*. Reidel.
van Benthem, J. (1989). Semantic parallels in natural language and computation. In H.-D. Ebbinghaus, J. Fernandez-Prida, M. Garrido, D. Lascar, & M. R. Artalejo (Eds.), *Logic colloquium, Granada 1987* (pp. 331–375). North Holland.
van Benthem, J. (1996). Modal logic as a theory of information. In J. Copeland (Ed.), *Logic and reality. Essays on the legacy of Arthur Prior* (pp. 135–168). Clarendon Press.
van Benthem, J. (2006). Logic in philosophy. In D. Jacquette (Ed.), *Handbook of the philosophy of logic* (pp. 65–99). Elsevier.
van Benthem, J. (2016). Tales from an old manuscript. In J. van Eijck, R. Iemhoff, & J. Joosten (Eds.), *Liber Amicorum Alberti* (pp. 5–14). College Publications.
van Benthem, J. (2018). Constructive agents. *Indagationes Mathematicae, 29*(1), 23–35.
van Benthem, J., & Bezhanishvili, G. (2007). Modal logics of space. *Handbook of spatial logics* (pp. 217–298). Springer.
van Benthem, J., & Bezhanishvili, N. (2021). Modern perspectives on filtration. In F. Faroldi & F. van der Putte (Eds.), *Kit Fine on modal logic*. Springer.
van Benthem, J., & Pearce, D. (1984). A mathematical characterization of interpretation between theories. *Studia Logica, 43*, 295–303.
van Benthem, J., Blackburn, P., & Wolter, F. (Eds.). (2006). *Handbook of modal logic*. Elsevier.
van Benthem, J. (2009). The information in intuitionistic logic. *Synthese, 167*(2), 251–270.
van Benthem, J. (2019). Implicit and explicit stances in logic. *Journal of Philosophical Logic, 48*, 571–601.
van Benthem, J., Bezhanishvili, N., & Holliday, W. (2017). A bimodal perspective on possibility semantics. *Journal of Logic and Computation, 27*, 1353–1389.
Veltman, F. (1984). Data semantics. In J. Groenendijk, T. Janssen, & M. Stokhof (Eds.), *Truth, interpretation, and information* (pp. 43–63). Foris.
Venema, Y. (1992). *Many-dimensional modal logic* (Unpublished doctoral dissertation). ILLC, University of Amsterdam.

Venema, Y. (1991). A modal logic for chopping intervals. *Journal of Logic and Computation, 1*(4), 453–476.
Vickers, S. (1989). *Topology via logic*. Cambridge University Press.
Wansing, H. (1993). *The logic of information structures*. Springer.
Wansing, H. (Ed.). (1996). *Proof theory of modal logic*. Springer.
Weiss, Y. (2021). A characteristic frame for positive intuitionistic and relevance logic. *Studia Logica, 109*, 687–699.
Wiener, N. (1914). A contribution to the theory of relative position. *Proceedings Cambridge Philosophical Society, 17*, 441–449.
Winskel, G. (1989). An introduction to event structures. In J. de Bakker, W.-P. de Roever, & G. Rozenberg (Eds.), *Linear time, branching time and partial order in logics and models for concurrency* (pp. 364–397). Springer.

# Chapter 6
# Entailment, Mingle and Binary Accessibility

**Katalin Bimbó and Jon Michael Dunn**

**Abstract** Saul Kripke's work on the semantics of modal logics is well known, unlike his work on Anderson and Belnap's system **E** of Entailment (a *modal relevance logic*), which included his proof of the decidability of its implicational fragment $\mathbf{E}_{\rightarrow}$, and also a counterexample to the conjecture of Belnap that $\mathbf{E}_{\rightarrow}$ is the intersection of the implicational fragments of the relevance logic **R** and the modal logic **S4**. This led to Storrs McCall's suggesting that the "mingle" axiom $A \rightarrow (A \rightarrow A)$ might be added to $\mathbf{E}_{\rightarrow}$, and that this system might be used instead of $\mathbf{S4}_{\rightarrow}$. We give a counterexample to this conjecture. We also disprove the conjecture Anderson and Belnap attribute to McCall, in which "restricted mingle" $\overrightarrow{A} \rightarrow (\overrightarrow{A} \rightarrow \overrightarrow{A})$ is added to $\mathbf{E}_{\rightarrow}$. We use a modal extension of the lattice $M_0$ and follow ideas of Meyer. In the 1970s, Dunn showed that **RM** (**R** with the mingle axiom) can be modeled using a binary accessibility relation, like the accessibility relation in Kripke's semantics for normal modal logics and his semantics for intuitionistic logic. Dunn also gave a variant interpretation for **EM** (**E** with the mingle axiom). We explore some other variations of these binary relational semantics.

**Keywords** Entailment · Hereditary condition · Mingle axiom · Modal logic · Relational semantics · Relevance logic · Trivalence

www.ualberta.ca/~bimbo.

K. Bimbó (✉)
Department of Philosophy, University of Alberta, 2–40 Assiniboia Hall, Edmonton T6G 2E7, Canada
e-mail: bimbo@ualberta.ca

J. M. Dunn
Luddy School of Informatics, Computing and Engineering, and Department of Philosophy, Indiana University, 901 East 10th Street, Bloomington, IN 47408–3912, USA

© The Author(s), under exclusive license to Springer Nature Switzerland AG 2024

Y. Weiss and R. Birman (eds.), *Saul Kripke on Modal Logic*, Outstanding Contributions to Logic 30, https://doi.org/10.1007/978-3-031-57635-5_6

## 6.1 Introduction

Saul Kripke has made a number of significant contributions to philosophy and to logic. Certainly the best known, at least among logicians, is his creation of the "possible world semantics" for modal logics in his papers (Kripke 1959a, 1963, 1965b), which is now often called the "Kripke semantics" for modal logic. This is not just a contribution to logic, but is set in a context where it also relates to philosophical issues in metaphysics, epistemology, and philosophy of language. The most striking aspect of this semantics is his use of a binary relation $R$ on worlds of "relative possibility," or "accessibility." But Kripke also made early contributions to relevance logic, including the proof that the implicational fragment of the system **E** of Entailment is *decidable*.[1] It thus seems appropriate for us to have a paper here that discusses variants of the system **E** of **E**ntailment and the system **R** of **R**elevant implication, both variants of which use a binary accessibility relation $R$. The first of these is modal, whereas the second is non-modal.

Another early link of Kripke to relevance logic came about through a conjecture of Anderson and Belnap. They say Anderson and Belnap (1975, p. 94): "The fact that $\mathbf{FE}_{\to}$ [a natural deduction system for the implicational fragment of the system **E** of entailment] arises by combining the subscript restrictions of $\mathbf{R}_{\to}$ with the reiteration restrictions of $\mathbf{FS4}_{\to}$ might lead one to think that $\mathbf{E}_{\to}$ contains as theorems all and only those formulas which are theorems of *both* $\mathbf{R}_{\to}$ and $\mathbf{S4}_{\to}$. We are indebted to Saul Kripke for a counterexample …."[2] The counterexample is

$$A \to [(A \to (A \to B)) \to (A \to B)],$$

which is not a theorem of $\mathbf{E}_{\to}$, because no implication is a theorem that has an implication as a consequent and a propositional variable as its antecedent (put in $p$ for $A$). This "Ackermann property," as it is called in Anderson and Belnap (1975), p. 40), was proved in Ackermann (1956).

This led Storrs McCall to suggest to JMD in a graduate course he took from McCall in 1964 that the implicational fragment $\mathbf{EM}_{\to}$ of the system **E**-**M**ingle (**E** with the mingle axiom) might be substituted for $\mathbf{S4}_{\to}$ in the conjecture. The mingle axiom $M$ is $A \to (A \to A)$. JMD was never able to prove this, but he did study various properties of the related system **RM** (**R** with the mingle axiom $M$). Anderson and Belnap (1975) seem to misunderstand, or perhaps, misremember McCall's suggestion, stating instead (p. 94) as a conjecture (attributed to McCall) that the system was not the implication fragment of **E** with the mingle axiom, but rather **E** with $\overrightarrow{M}$, the

[1] It is remarkable that Kripke started to work on modal and relevance logic while still a high school student. The formulation of the sequent calculus $LE_{\to}$ together with the proof of its decidability date to the summer and fall of 1959.

[2] The conjecture was originally formulated in Belnap's Ph.D. dissertation, Belnap (1959).

restricted form of the mingle axiom $(A \to B) \to [(A \to B) \to (A \to B)]$ added.[3] Let's call that system $\mathbf{E}\overrightarrow{\mathbf{M}}$. Both conjectures are interesting, and it is worth pointing out that the conjecture that $\mathbf{E}_{\to} = \mathbf{R}_{\to} \cap \mathbf{EM}_{\to}$ can be motivated by replacing the principal type of $\mathsf{K}$ (in $\mathbf{S4}_{\to}$) by a special instance of it. $\mathbf{E}\overrightarrow{\mathbf{M}}_{\to}$ further limits the latter formula in a way that avoids violating the Ackermann property.

The *Kripke-semantics for modal logic* has been a game-changer and has led to applications not just to metaphysics but also epistemology, tense logic, deontic logic, program verification, etc. An application important for our work here was Kripke's (1965a) semantics for intuitionistic logic (a semantics that was independently discovered by Grzegorczyk, 1964). The basic idea of this semantics for intuitionistic logic is to have "evidential situations" in place of the "possible worlds" in Kripke's semantics for modal logic, and to understand the binary relation not as "relative possibility," but something like one evidential situation is included in another. It is required that if an atomic sentence is true in an evidential situation that is included in a second, then that sentence stays true in the second. This "Hereditary Property" can then be proved by induction to hold for compound sentences as well.

## 6.2 A Kripke-Style Semantics for R-Mingle

There was a time though when relevance logic and other so-called substructural logics did not have a "Kripke-style semantics." But in the early 1970s, Robert K. Meyer and Richard Routley produced a semantics for the systems **E** and **R** and related logics based on a *ternary accessibility relation* (3-placed, not 2-placed).[4]

Routley and Meyer were able to provide a complete *ternary* semantics for **RM** too, but Dunn (1976a) was able to provide a complete *binary* semantics for **RM** as a modification of Kripke's and Grzegorczyk's semantics for intuitionistic logic. The basic modification made by Dunn was to allow a sentence to be both true and false (but not neither), and to require a *trivalent hereditary condition* that requires falsity preservation as well as truth preservation. If we view "situations" as pieces of information ("information states"), both of these requirements are natural (at least if we stay away from non-monotonic increases). This is much easier to understand that the Routley–Meyer semantics for **RM** with its ternary accessibility relation. For a recent attempt to give informal interpretations to the ternary relation, see Beall et al. (2012). For an axiomatization of **RM**, see the Appendix.

One negative feature of the semantics for **RM**, as we will see below, is that it satisfies the *chain theorem* $(A \to B) \vee (B \to A)$, which requires that the information

[3] A piece of evidence to support JMD's memory that McCall intended the non-restricted form of mingle is McCall's (1963) Ph.D. dissertation, p. 158, in which he discusses adding a rule equivalent to the unrestricted mingle axiom to Kripke's Gentzen system for $\mathbf{E}_{\to}$ (which McCall denotes as $LI$).

[4] See Bimbó and Dunn (2017) for a history of how a number of other logicians came up with equivalent or closely related semantics about the same time. However, the name "Routley–Meyer semantics" had the same sticking quality as "Kripke semantics."

states are linearly ordered. This is unlike the semantics of intuitionistic logic, and reflects that **RM**'s characteristic matrices are Sugihara matrices, which are chains. It was shown in Dunn (1976c) that one can give a Kripke-style semantics for **EM** by simultaneously dropping both the requirement that the accessibility relation be connected and the hereditary condition. In the present paper, we will review these earlier results and consider some further variations. Since the result about **EM** was only "semi-published" in the *Relevance Logic Newsletter*, this also gives us an opportunity to present it to a wider audience and to give it a more permanent venue.[5] Additionally, it is the first step in exploring other possible modifications. You might view this paper as a "voyage of discovery." We write it in a style that doesn't just prove theorems as quickly as possible, but rather explores these proofs and how they might, or might not, be modified.

**Definition 6.2.1** An **RM** *model structure* is an ordered-triple $\langle G, K, R\rangle$, where $K$ is a non-empty set, $G \in K$, and $R$ is a chain, i.e., a total ordering. That is, $R$ is a binary relation on $K$, which is reflexive, anti-symmetric, transitive and connected (i.e., for $H, H' \in K$, $HRH'$ or $H'RH$ holds).[6] We further require that $G$ be the least element in the sense that for all $H \in K$, $GRH$.

An **RM** *model* adds a *valuation* $v$, which assigns a non-empty set of the truth values $t$ and $f$ to each atomic sentence $p$ at each state $H$, i.e., $\emptyset \neq v(p, H) \subseteq \{t, f\}$. We call such a valuation *trivalent*. A valuation has to satisfy the following trivalent hereditary condition.

**Definition 6.2.2** (*Trivalent Hereditary Condition*) For each atomic sentence $p$, if $HRH'$ then $v(p, H) \subseteq v(p, H')$, i.e., if $t \in v(p, H)$ and $HRH'$ then $t \in v(p, H')$, and if $f \in v(p, H)$ and $HRH'$ then $f \in v(p, H')$.

Note that the Kripke–Grzegorczyk semantics for intuitionistic logic has a similar hereditary condition that is simply truth preserving, whereas the condition above is falsity preserving as well. This reflects taking an even-handed position to truth and falsity. Also, our trivalent hereditary condition respects the idea that an information state might simultaneously say that a sentence is true and that the sentence is false. We do not allow though an information state to be indecisive about the truth or the falsity of a sentence; we shall consider this option later on.

Although the hereditary condition is stated only for atomic sentences, it holds for sentences generally. It can be shown once we state additional requirements on how a valuation $v$ can be extended to compound sentences.

---

[5] The *Newsletter* was distributed for less than two years (1975–1976), and was first edited by Kenneth Collier and Robert Wolfe, and then by just Robert Wolfe. It was privately distributed to a list and consisted of "xeroxed" typewritten pages. It currently is available on the Website of the Australasian Association for Logic at the URL `sites.google.com/view/aalogic/curios/relevance-logic-newsletter` (as PDF files of scans). (Accessed on June 3rd, 2024.)

[6] Anti-symmetry is optional, the proofs work with or without it.

**Definition 6.2.3** A valuation $v$ is extended to all formulas according to the following clauses.

$(\sim t)$ $t \in v(\sim A, H)$ iff $f \in v(A, H)$;
$(\sim f)$ $f \in v(\sim A, H)$ iff $t \in v(A, H)$;
$(\wedge t)$ $t \in v(A \wedge B, H)$ iff $t \in v(A, H)$ and $t \in v(B, H)$;
$(\wedge f)$ $f \in v(A \wedge B, H)$ iff $f \in v(A, H)$ or $f \in v(B, H)$;
$(\vee t)$ $t \in v(A \vee B, H)$ iff $t \in v(A, H)$ or $t \in v(B, H)$;
$(\vee f)$ $f \in v(A \vee B, H)$ iff $f \in v(A, H)$ and $f \in v(B, H)$;
$(\rightarrow t)$ $t \in v(A \rightarrow B, H)$ iff for all $H'$ such that $HRH'$, $t \in v(A, H')$ only if $t \in v(B, H')$, and $f \in v(B, H')$ only if $f \in v(A, H')$;
$(\rightarrow f)$ $f \in v(A \rightarrow B, H)$ iff (counterexample) $t \in v(A, H)$ and $f \in v(B, H)$, or (non-true) $t \notin v(A \rightarrow B, H)$.

**Lemma 6.2.4** *An* **RM** *model has the* trivalent hereditary property*, i.e., for any A, if* $HRH'$ *then* $v(A, H) \subseteq v(A, H')$.

***Proof*** Consider an **RM** model $\langle G, K, R, v\rangle$. We now prove the trivalent hereditary property, that is, the information order $R$ preserves falsity as well as truth. By definition, the valuation has this property when applied to atomic sentences. Then the proof proceeds straightforwardly, using formula induction, until one gets to the case $(\rightarrow f)$. (We omit the easy cases.) If $f \in v(A \rightarrow B, H)$, because $t \in v(A, H)$ while $f \in v(B, H)$, then all is well, since then for any $H'$ such that $HRH'$ we can obtain by inductive hypothesis that $t \in v(A, H')$ while $f \in v(B, H')$, and so $f \in v(A \rightarrow B, H')$. But what happens if $f \in v(A \rightarrow B, H)$ because $t \notin v(A \rightarrow B, H)$? We have to consider two cases.

(a) $\exists H_1(HRH_1$ such that $t \in v(A, H_1)$ and $t \notin v(B, H_1))$;
(b) $\exists H_1(HRH_1$ such that $f \in v(B, H_1)$ and $f \notin v(A, H_1))$.

The cases are fully symmetric, so we shall just consider (a). We have assumed $HRH_1$. But again there are two cases: (i) $H_1RH'$ or (ii) $H'RH_1$. Let us start with case (i). So we have $HRH_1RH'$, $t \in v(A, H_1)$ and $t \notin v(B, H_1)$. Since $t \notin v(B, H_1)$, then $f \in v(B, H_1)$. So by inductive hypothesis, since $t \in v(A, H_1)$ and $f \in v(B, H_1)$ and $H_1RH'$, then $t \in v(A, H')$ and $f \in v(B, H')$. So $f \in v(A \rightarrow B, H')$.

In case (ii), we have $HRH'RH_1$. Again we are assuming $t \notin v(A \rightarrow B, H)$ and that this is because $t \in v(A, H_1)$ and $t \notin v(B, H_1)$. So $t \notin v(A \rightarrow B, H')$ since $H'RH_1$. But then $f \in v(A \rightarrow B, H')$. □

Now we turn to specifying a particular model, the so-called canonical model.

**Definition 6.2.5** A *theory* $T$ is a set of sentences closed under provable implication (i.e., if $A \in T$, and $A \rightarrow B$ is a theorem of **RM**, then $B \in T$), and adjunction (i.e., if $A, B \in T$, then $A \wedge B \in T$). A *regular theory* is one that contains all of the theorems of **RM**. A *prime theory* is such that if $A \vee B \in T$, then at least one of $A$ or $B \in T$.

**Definition 6.2.6** A *canonical* **RM** *model structure* is an ordered triple $\langle G, K, R\rangle$, where $G$ is a prime regular theory, $K$ is the a set of prime regular theories extending $G$, and $R$ is $\subseteq$ (restricted to $K$).

Clearly $R$ is a reflexive, anti-symmetric and transitive relation on $K$, and $G$ is the least member of the set $K$. It is easy to prove that, in addition, $R$ is a total order. Thus, suppose to the contrary that we have two prime regular theories $T$ and $T'$ such that $T \subsetneq T'$ and $T' \subsetneq T$. Then there is a sentence $A \in T$ such that $A \notin T'$ and a sentence $B$ such that $B \in T'$ and $B \notin T$. $(A \to B) \vee (B \to A)$ is a theorem of **RM**, so $(A \to B) \vee (B \to A) \in G$ (since $G$ is regular). And since $G$ is prime, either $A \to B \in G$ or $B \to A \in G$. If the first then $B \in T$ (contradiction!), and if the second then $A \in T'$ (contradiction!).[7]

**Definition 6.2.7** A *canonical* **RM** *model* adds a *canonical valuation* $v$, which assigns a non-empty set of the truth values $t$ and $f$ to each atomic sentence $p$ at each state $H$, so that $t \in v(p, H)$ iff $p \in H$, and $f \in v(p, H)$ iff $\sim p \in H$.

We would say that all the clauses in Definition 6.2.3 are natural for **RM** with the possible exception of $(\to f)$. We would like to leave off the "escape clause," that is, "or $t \notin v(A \to B, H)$." But it seems this disjunctive requirement is needed to keep the valuations trivalent. *Without* it, an implicational sentence $A \to B$ could end up with neither $t$ nor $f \in v(A \to B, H)$. *With* the requirement, it is easy to show trivalence by an induction on formulas. (We only state this as a lemma and leave the easy induction to the reader.)

**Lemma 6.2.8** *If for every state $H$ and for each atomic sentence $p$, $t \in v(p, H)$ or $f \in v(p, H)$, then the same holds for* compound sentences *as well.*

The following then is immediate.

**Theorem 6.2.9** (Trivalence) *For every sentence $A$, $t \in v(A, H)$ or $f \in v(A, H)$.*

We next prove a lemma linking the truth and falsity of a sentence in the canonical model not just to the provability of atomic sentences and their negations in prime theories, but to all sentences.

**Theorem 6.2.10** (Canonical Equivalence) *For every sentence $A$, the following "canonical equivalences" hold for the canonical model $\langle G, K, R, v\rangle$.*

$(t)$ $t \in v(A, H)$ *iff* $A \in H$;
$(f)$ $f \in v(A, H)$ *iff* $\sim A \in H$.

***Proof*** Again the proof is by formula induction. The base case is by definition, and the other cases, except for one, are straightforward. The latter includes the case $t \in v(A \to B, H)$ if $A \to B \in H$, and its contraposed converse $t \notin v(A \to B, H)$ if $A \to B \notin H$; we shall assume that these have been proved. We now examine the exceptional case, $f \in v(A \to B, H)$.

The clause $(\to f)$ says $f \in v(A \to B, H)$ iff either $t \in v(A, H)$ and $f \in v(B, H)$, or $t \notin v(A \to B, H)$. By inductive hypothesis, the right-hand side can be simplified to $A \in H$ and $\sim B \in H$, or $t \notin v(A \to B, H)$. And as we said in the previous paragraph, this last disjunct can be further simplified to $A \to B \notin H$. So we

[7] We are not "jumping to conclusions" here. $(C \wedge (C \to D)) \to D$ is a theorem of **R**, hence, of **RM**.

have $f \in v(A \to B, H)$ iff either $A \in H$ and $\sim B \in H$, or $A \to B \notin H$. Now we can reason simply about formulas and the prime regular theory $H$ to show $A \in H$ and $\sim B \in H$, or $A \to B \notin H$ iff $\sim(A \to B) \in H$.

For left-to-right, we have to consider two cases. (1) $A \in H$, $\sim B \in H$, and (2) $A \to B \notin H$. In case (1), since $H$ is a theory it is closed under adjunction, and so $A \wedge \sim B \in H$. But it is a theorem of **RM** that $(A \wedge \sim B) \to \sim(A \to B)$, and theories are closed under provable implications. Thus, $\sim(A \to B) \in H$. In case (2), since $H$ is a regular theory, $(A \to B) \vee \sim(A \to B) \in H$, and since $H$ is prime and $A \to B \notin H$, then $\sim(A \to B) \in H$. So we get our desired conclusion $\sim(A \to B) \in H$ in either case.

For right-to-left, we assume that $\sim(A \to B) \in H$. We need to show that $A \in H$ and $\sim B \in H$, or else $A \to B \notin H$. Let us suppose the last disjunct is false, i.e., that $A \to B \in H$, and then show that the first disjunct is true, by showing $A \wedge \sim B \in H$.[8]

The formula $(A \to B) \to [(\sim A \vee B) \to (A \to B)]$ is a theorem of **RM** (T11 in Dunn, 1970). Using the rules contraposition, De Morgan and double negation, each of which is derivable in **RM**, we can obtain $(A \to B) \to [\sim(A \to B) \to (A \wedge \sim B)]$. Since we are now assuming that both $A \to B, \sim(A \to B) \in H$, by two uses of modus ponens, we obtain $A \wedge \sim B \in H$. And with two uses of conjunction elimination, this can be simplified to $A, \sim B \in H$; hence, we are done. □

**Definition 6.2.11** Let $A$ be a sentence and $\Gamma$ a set of sentences. $A$ is *deducible* from $\Gamma$ in **RM** (in symbols, $\Gamma \vdash_{\mathbf{RM}} A$) iff there is a finite sequence of sentences $A_1, \ldots, A_i, A_{i+1}, \ldots, A_n$, where $A_n = A$, and each sentence $A_i$ is either a member of $\Gamma$, or an axiom of **RM**, or follows from previous members of the sequence by the rule of modus ponens, (i.e., $B, B \to C \Diamond C$) or by the rule of adjunction (i.e., $B, C \Diamond B \wedge C$).

There is a related notion of $A$ being an **RM** *logical consequence* of $\Gamma$ (in symbols $\Gamma \vDash_{\mathbf{RM}} A$) in the sense that for every **RM** model $\langle G, K, R, v\rangle$, if $t \in v(C, G)$ for every sentence $C \in \Gamma$, then $t \in v(A, G)$.

**Theorem 6.2.12** (Strong Completeness) *If* $\Gamma \vDash_{\mathbf{RM}} A$, *then* $\Gamma \vdash_{\mathbf{RM}} A$.

***Proof*** As usual, we prove the contrapositive. Suppose $A$ is not **RM** deducible from $\Gamma$. For a standard construction in Henkin-style proofs of completeness, we enumerate all of the sentences in the language of **RM** as $B_1, B_2, \ldots, B_n, B_{n+1}, \ldots$. Let $\Gamma_0 = \Gamma$, and let us define $\Gamma_{n+1} = \Gamma_n \cup \{B_{n+1}\}$ if $\Gamma_n \cup \{B_{n+1}\} \nvdash_{\mathbf{RM}} A$, and otherwise, $\Gamma_{n+1} = \Gamma_n$. Define $G = \bigcup_{n\in\mathbb{N}} \Gamma_n$. It is easy to show that $G$ is a regular prime theory and that $A \notin G$. Consider the canonical model structure $\langle G, K, R\rangle$. By the Canonical Equivalence Theorem 6.2.10, $t \in v(A, G)$ iff $A \in G$. But since $A \notin G$, $t \notin v(A, G)$, and yet for every $C \in \Gamma$, $C \in G$, i.e., $t \in v(C, G)$. □

**Theorem 6.2.13** (Strong Soundness) *If* $\Gamma \vdash_{\mathbf{RM}} A$, *then* $\Gamma \vDash_{\mathbf{RM}} A$.

[8] Any similarity that this shows to Disjunctive Syllogism is entirely acknowledged by us. Attempts to use relevance logics as metalogics for themselves seem not to work, as was shown by deep investigations by Meyer (1985). (See also Anderson et al., 1992, Sect. 80).

***Proof*** We remark that both of the rules of **RM** used in defining deducibility, namely modus ponens and adjunction, clearly preserve truth. It only needs to be verified that the axioms of **RM** are true at $G$ in every **RM** model $\langle G, K, R, v\rangle$. (We leave most of this to the reader, but we will check two of the axioms after stating a pair of corollaries and proving a useful lemma.) □

So by conjoining the last two theorems, we have a semantic characterization of deducibility in **RM**.

**Corollary 6.2.14** (Strong Soundness and Completeness) $\Gamma \vdash_{\mathbf{RM}} A$ *iff* $\Gamma \vDash_{\mathbf{RM}} A$.

**Corollary 6.2.15** (Weak Soundness and Completeness) $\vdash_{\mathbf{RM}} A$ *iff* $\vDash_{\mathbf{RM}} A$.

In this last statement, we omitted displaying the empty set (as usual). Since there are no premises, $\vdash_{\mathbf{RM}} A$ means that $A$ is a theorem of **RM**, and $\vDash_{\mathbf{RM}} A$ means that $A$ is **RM** valid.

Note that the logical consequence relation (i.e., $\vDash_{\mathbf{RM}}$) requires only preservation of $t$; it does not require "co-preservation" of $f$. This is unlike the semantical clause for implication ($\rightarrow t$). Let us then define a stronger notion of entailment.

**Definition 6.2.16** $\Gamma$ **RM** *entails* $A$ iff for every **RM** model $\langle G, K, R, v\rangle$, if $t \in v(C, G)$ for every sentence $C \in \Gamma$, then $t \in v(A, G)$, and also if $f \in v(A, G)$, then for some $C \in \Gamma$, $f \in v(C, G)$.

As a special case when $\Gamma = \{ C \}$, we have $\{ C \}$ **RM** entails $A$, which we write simply as $C$ **RM** entails $A$. This is useful in verifying the axioms.

**Lemma 6.2.17** (Entailment–Implication Lemma) *An implication* $A \rightarrow B$ *is* valid *in the class of* **RM** *models, i.e., for all* **RM** *models* $\langle G, K, R, v\rangle$, $t \in v(A \rightarrow B, G)$ *iff* $A$ **RM** entails $B$.

***Proof*** (*Left-to-right.*) Assume that for all **RM** models $\langle G, K, R, v\rangle$, $t \in v(A \rightarrow B, G)$. Then consider an arbitrary **RM** model $\langle G, K, R, v\rangle$. Since $t \in v(A \rightarrow B, G)$, then by ($\rightarrow t$), for all $H$ such that $GRH$, $t \in v(A, H)$ only if $t \in v(B, H)$, and $f \in v(B, H)$ only if $f \in v(A, H)$. This is exactly the definition of $A$ **RM** entails $B$.

(Right-to-left.) We assume $A$ **RM** entails $B$ and show that $A \rightarrow B$ is valid in the class of **RM** models. For the latter, we must prove that $t \in v(A \rightarrow B, G)$, for an arbitrary **RM** model $(G, K, R, v)$. That is, for any $H$ such that $GRH$, if $t \in v(A, H)$ then $t \in v(B, H)$, and if $f \in v(B, H)$ then $f \in v(A, H)$. But since we are assuming that $A$ **RM** entails $B$, we have that both of these hold for all information states $H \in K$, because $G$ is the least element of $K$ along the relation $R$. □

As in Dunn (1976a) (and as we promised above) we leave most of the work of checking the validity of the axioms to the reader. We publicly check two of the most central axioms.

***Proof (of Theorem 6.2.13 continued)*** Obviously, we should examine the *mingle axiom* $A \to (A \to A)$. By the Entailment–Implication Lemma, it suffices to suppose $t \in (A, G)$, and then show that $t \in (A \to A, G)$, and then, to prove that falsity is preserved in the opposite direction. The first is obvious, since for any $H$ such that $GRH$, if $t \in (A, H)$ then $t \in (A, H)$. That's easy, but wait, we are not finished yet; we must still prove falsity preservation in the reverse direction. So suppose that $f \in (A \to A, G)$. Then either $t \notin (A \to A, G)$, or $t \in (A, G)$ while $f \in (A, G)$. But the first cannot hold, and if the second holds, then $f \in (A, G)$ as desired.

The other axiom we want to verify is a typical axiom of the system **R** of relevant implication that distinguishes it from the modal system **E** of entailment, namely,

$$A \to [(A \to B) \to B] \qquad \text{(assertion)}.$$

It is important to note that the Trivalent Hereditary Lemma (Lemma 6.2.4) is critical to the validity of assertion. We will **emphasize** this in the proof below.

Again we use the Entailment–Implication Lemma. We first assume that $t \in v(A, H)$ and show $t \in v((A \to B) \to B, H)$. There are two parts to the latter: (i) we must show that if $HRH'$ and $t \in v(A \to B, H')$ then $t \in v(B, H')$, and (ii) if $HRH'$ and $f \in v(B, H')$ then $f \in v(A \to B, H')$. So let's stipulate $HRH'$.

For (i), we suppose $t \in v(A \to B, H')$ in order to show that $t \in v(B, H')$. But since $t \in v(A \to B, H')$ then for all $H''$ such that $H'RH''$, if $t \in (A, H'')$ then $t \in (B, H'')$. Since $R$ is reflexive, $H'$ counts as such an $H''$. Therefore, if $t \in (A, H')$ then $t \in (B, H')$. But we first assumed that $t \in v(A, H)$, and so by the **Trivalent Hereditary Lemma**, since $HRH'$, $t \in v(A, H')$, and then $t \in v(B, H')$, as desired.

For (ii), we assume that $f \in v(B, H')$ in order to show $f \in v(A \to B, H')$. Since we are assuming $t \in v(A, H)$, again by the **Trivalent Hereditary Lemma** we have $t \in (A, H')$. This, with our assumption that $f \in v(B, H')$, assures by $(\to f)$ that $f \in v(A \to B, H')$.

We still have to prove that assertion $(A \to [(A \to B) \to B])$ co-preserves $f$. So let us assume $f \in v((A \to B) \to B, H)$ and show $f \in v(A, H)$. According to $(\to f)$, there are two ways that we might have $f \in v((A \to B) \to B, H)$, either there is a counterexample (i.e., $t \in v(A \to B, H)$ and $f \in v(B, H)$), or $(A \to B) \to B$ is simply not true at $H$ (i.e., $t \notin v((A \to B) \to B, H)$). If the former, then since $t \in v(A \to B, H)$, $HRH$, and $f \in v(B, H)$, then $f \in v(A, H)$, as desired. So let us consider the latter way, $t \notin v((A \to B) \to B, H)$. This will get a little tricky! You might (as once we did) utilize the Trivalence Theorem (Theorem 6.2.9) and say that $f \in v((A \to B) \to B, H)$ and that there are two ways in which $f \in v((A \to B) \to B, H)$, etc. But on that path we get stuck. Then we remember that we had already verified that assertion is truth preserving. So since $t \notin v((A \to B) \to B, H)$, $t \notin v(A, H)$. But then, by the Trivalence Theorem, $f \in (A, H)$, as required. □

## 6.3 Pivotal Points in the Semantics for RM

We shall discuss certain core features of this semantics. There are a couple of requirements stipulated in the Kripke-style semantics for **RM** that we might consider playing with, namely, the *trivalent hereditary condition*, the *trivalent information order*, the *chain theorem* and *trivalence*. We shall scrutinize each one after another, and in the process we will find that they turn out to be tangled together in interesting ways.

**1. The Trivalent Hereditary Condition.** This condition is quite natural if we think of the relation $R$ as an information order. Something that distinguishes the information order for **RM** from the information order for intuitionistic logic is that the latter requires only truth preservation whereas the former also requires falsity preservation.[9] This is an even-handed treatment of truth and falsity, whereas intuitionistic logic can be interpreted as favoring truth, in a constructive sense of provability. This Trivalent Hereditary Condition is unsurprising if we think of $R$ as an information order, where as information increases, some sentences are shown true and some sentences are shown false. Of course, some sentences are shown to be both, but this is part of the whole point of **RM** being a paraconsistent logic (see Dunn, 2021). However, the Trivalent Hereditary Condition is not natural at all if we think of $R$ as some kind of modal accessibility relation.

**2. Conditions on the Trivalent Information Order.** There are several things to observe about the *trivalent information order*. If we think of $R$ this way (i.e., $HRH'$ means that all the information in state $H$ is contained in state $H'$), then it is straightforward to take $R$ be reflexive and transitive (even anti-symmetric). However, it seems very artificial to require that $R$ be also a linear order, for that requires the addition of *connectedness* (i.e., $HRH'$ or $H'RH$). Why do we require connectedness? It has to do with the Trivalent Hereditary Property. There appears to be no way to prove it without connectedness. (See the proof of the trivalent hereditary property in Lemma 6.2.4.)

**3. Chain Theorem.** It is connectedness of course that leads to the validity of the Chain Theorem $(A \to B) \vee (B \to A)$, which is provable in **RM**. This is a very awkward theorem, particularly, for a relative of a relevance logic. It says that given any two sentences, even though highly irrelevant, at least one of them implies the other. For example, either the moon is made of green cheese implies that cats meow, or cats meow implies that the moon is made of green cheese.

**4. Trivalence.** Another unnatural requirement is that $v(p, H)$ is non-empty. Why can't the information state $H$ have nothing to say about $p$? Given the natural 4-valued models for the system **FDE** (True, False, Both, and Neither) dropping the requirement of trivalence seems quite natural. Besides the clause "or $t \notin v(A \to B, H)$" in $(\to f)$ seems ad hoc.

---

[9] Dretske (1981) and Floridi (2011) have each argued that information must be true. Fetzer (2004) and Dunn have argued against this "veridicality thesis." See Dunn (2013) for a discussion. Here, not surprisingly, we side with Dunn.

To our minds a very reasonable variation would be to drop the requirement that $R$ be connected. Then allow $v(p, H)$ to be empty, and in concert with this, drop the escape clause "or $t \notin v(A \to B, H)$" from $(\to f)$. That clause would then read as $f \in v(A \to B, H)$ iff $t \in v(A, H)$ and $f \in v(B, H)$, which is straightforward in a 4-valued setting. $A \to B$ is true according to an information state iff at all increased states of information, if $A$ is true then $B$ is true, and if $B$ is false then $A$ is false $(\to t)$, and $A \to B$ is false according to an information state iff $A$ is true and that $B$ is false according to that information state $(\to f)$. This allows $A \to B$ to be true, false, both, or neither at a given information state.

The problem though is that the Canonical Equivalence Theorem (Theorem 6.2.10) says for the case $(f)$ $f \in v(A \to B, H)$ iff $\sim(A \to B) \in H$. The left-to-right direction is OK: $f \in v(A \to B, H)$ implies $t \in v(A, H)$ and $f \in v(B, H)$, and by inductive hypothesis, these last two are equivalent to $A \in H$ and $\sim B \in H$. From these, we get $A \wedge \sim B \in H$, and then $\sim(A \to B) \in H$, just as we did above. But from right to left there is a problem. Let us assume $\sim(A \to B) \in H$. We would somehow have to prove that $t \in v(A, H)$ and $f \in v(B, H)$, i.e., by inductive hypothesis that $A \in H$ and $\sim B \in H$, but this does not seem possible.

Let us consider the number of logics that might—in principle—be investigated.

- Connectedness of the accessibility relation $R$ can hold or not hold (2).
- The Hereditary Condition can preserve $t$, co-preserve $f$, or both, or we can not have the Hereditary Condition (4).
- The Valuation Condition for $\to$ can require preservation of $t$, co-preservation of $f$, or both (3).
- Valuations can assign a sentence Both $t$ and $f$ or Neither $t$ nor $f$, or both (3).
- Validity is defined at $t$-states or at all states (2).
- Logical consequence preserves $t$, co-preserves $f$, or both (3).

Doing the math, we calculate that in principle $2 \times 4 \times 3 \times 3 \times 2 \times 3 = 432$ different logics that we might investigate. Now many of these have little or no intuitive motivation or application. Why would one, for example, investigate a hereditary condition that preserves only falsity, or an implication that does not preserve truth but co-preserves falsity, or a logical consequence relation that doesn't preserve truth but co-preserves falsity? And with others it seems that the conditions just do not fit well together for making a logic of that type.

Regrettably, we seem to find that dropping the Connectedness Condition while keeping the strongest version of the Hereditary Condition (preservation of both $t$ and $f$) appears to be an example of this.

It is difficult to avoid a rather direct proof of the Chain Theorem—using only axioms and rules of **R**—from the **RM** theorem Implication–Negation Safety:[10]

$$\sim(A \to A) \to (B \to B).$$

[10] Meyer (1968) proved a more general form, noting that if $\vdash_{\mathbf{RM}} \sim A$ and $\vdash_{\mathbf{RM}} B$, then $\vdash_{\mathbf{RM}} A \to B$. The label "Safety" relates to the fact that even if one has the negation of a theorem, it does not imply everything but only what is already a theorem.

We sketch proofs that, indeed, use only axioms of **R**.
First, we can prove *Ex Falso Verum* $\sim A \to (B \to (A \to B))$ from Implication–Negation Safety.

1. $\sim(B \to B) \to (A \to A)$ Implication–Negation Safety
2. $A \to (\sim(B \to B) \to A)$ 1, Permutation
3. $A \to (\sim A \to (B \to B))$ 2, Contraposition and Double Negation
4. $\sim A \to (A \to (B \to B))$ 3, Permutation
5. $\sim A \to (B \to (A \to B))$ 4, Permutation.

Next, we will prove the Strong Chain Theorem $\sim(A \to B) \to (B \to A)$.

1. $\sim A \to (B \to (A \to B))$ *Ex Falso Verum*
2. $\sim A \to (\sim(A \to B) \to \sim B)$ 1, Contraposition
3. $\sim(A \to B) \to (\sim A \to \sim B)$ 2, Permutation
4. $\sim(A \to B) \to (B \to A)$ 3, Contraposition.

Given that $(\sim C \to D) \to (C \vee D)$ is a well-known theorem of **R** (and of **E**), hence, also of **RM**, the Chain Theorem follows from the Strong Chain Theorem. (It is sufficient to take $A \to B$ for $C$ and $B \to A$ for $D$.)

The proofs above are due to Robert K. Meyer (1968) and use only axioms of **R**, except of course for $\sim(B \to B) \to (A \to A)$. But this last follows fairly easily from the Mingle axiom $C \to (C \to C)$ given the presence of conjunction, again shown by Robert K. Meyer. Moreover, that $\sim(B \to B) \to (A \to A)$ is valid follows straightforwardly given trivalence and the semantic definition of $\to$ as both truth preserving and falsity co-preserving, since its consequent is always true and its antecedent is always false.

So how can we avoid it? *One way* is "to go modal," and to strengthen the implication of **R** into the entailment of **E**, which combines necessity and relevance. Permutation is used a number of times in the proofs above, and Permutation is invalid in **S4** and its relevant version **E**. Instead, these modal systems tolerate a special instance of Permutation:

$$[A \to (\overrightarrow{B} \to C)] \to [\overrightarrow{B} \to (A \to C)] \quad \text{Restricted Permutation,}$$

where the arrow on top of the $B$ indicates that it must be of the form $B \to D$. (Toward the end of the next section we shall mention another connection between modality and the absence of full permutation.)

*Another way* might be to consider valuations that are not trivalent, but instead depart from classical valuations by not just allowing a sentence to receive both of the values $t$ and $f$, but also neither of the truth values $t$ and $f$.

We explore each of these two options, one after the other, in the following two sections.

## 6.4 A Four-Valued Version of the Kripke-Style Semantics for RM

The first thought we might have is that one problem caused by allowing a sentence to be assigned the empty set of truth values is that it is then too easy for at least some theorems to end up not true. This is certainly the case given our valuation clauses for **RM** above. Suppose $v(p, H) = \emptyset$. Then $v(p \vee \sim p, H) = \emptyset$, and this is true even when $H = G$. Yet $p \vee \sim p$ is a theorem of **RM**. This does not hold for all theorems, consider $A \rightarrow A$; still, it is problematic. A way out is to require at least $G$ (or maybe, even some larger subset of information states) to be *normal*, i.e., not to assign an empty set of truth values to an atomic sentence, and then require for validity just truth at normal states. (This is in effect what the Routley–Meyer ternary semantics does for **RM**, and similarly, for **R** and **E**.)

We should also modify the clause $(\rightarrow f)$, that is,

$(\rightarrow f)$ $f \in v(A \rightarrow B, H)$ iff either (counterexample) $t \in v(A, H)$ and $f \in v(B, H)$, or (non-true) $t \notin v(A \rightarrow B, H)$.

By dropping the "non-true" escape clause, we get

$(\rightarrow f')$ $f \in v(A \rightarrow B, H)$ iff (counterexample) $t \in v(A, H)$ and $f \in v(B, H)$.

Otherwise, implications would be either true or false (including both), whereas formulas that are not implications could sometimes be neither. This modification gives a side benefit. It was this escape clause that was getting in the way of proving the Trivalent Hereditary Lemma without requiring that the accessibility relation be connected. Now there is no problem with the Quadrivalent Hereditary Condition, as the reader can easily see.

But there is a problem with proving the Canonical Equivalence Theorem, in particular, with a case in clause $(f)$, when the formula is an implication. That is,

$$f \in v(A \rightarrow B, H) \quad \text{iff} \quad \sim(A \rightarrow B) \in H.$$

There is no problem with the left-to-right direction. The proof proceeds as before except that there is no reason now to consider the case $t \notin v(A \rightarrow B, H)$. We might assume $f \in v(A \rightarrow B, H)$ and conclude using $(\rightarrow f')$ that $t \in v(A, H)$ and $f \in v(B, H)$. By inductive hypothesis, these would be equivalent to $A \in H$ and $\sim B \in H$, and since $H$ is closed under adjunction and $(A \wedge \sim B) \rightarrow \sim(A \rightarrow B)$ is a theorem of **RM** (indeed, of **E**), then $\sim(A \rightarrow B) \in H$.

But to show the right-to-left direction looks impossible. Suppose we assume $\sim(A \rightarrow B) \in H$. The only way to get to $f \in v(A \rightarrow B, H)$ seems to be via establishing that $t \in v(A, H)$ and $f \in v(B, H)$. We might try to obtain these by getting $A \in H$ and $\sim B \in H$, and then using the induction hypothesis. However, there is no way to get from our assumption $\sim(A \rightarrow B) \in H$ to $A \in H$ and $\sim B \in H$. We do not have as **RM** theorems $\sim(A \rightarrow B) \rightarrow A$ or $\sim(A \rightarrow B) \rightarrow \sim B$. From the former, by Contraposition and Double Negation, we would get $\sim A \rightarrow (A \rightarrow B)$,

and then $(A \wedge {\sim}A) \rightarrow B$, which is well known not to be a theorem of **RM**. Also, $B \rightarrow (A \rightarrow B)$ is immediate. (Any of the last three formulas can easily be shown not to be a theorem of **RM** using the 3-element "Sugihara matrix." See Meyer (1968) for details.)

Dunn (2000) studies this system and names it $\mathbf{BN}_{1,0}$ to indicate that it allows *B*oth of the truth values (i.e., $\{t, f\}$) and *N*either of them (i.e., $\emptyset$), and also requires that $\rightarrow$ preserve both 1 (here $t$) forward and 0 (here $f$) backwards.[11] He also studies $\mathbf{BN}_1$, $\mathbf{B}_{1,0}$, $\mathbf{B}_1$, $\mathbf{N}_{1,0}$, $\mathbf{N}_1$, and then, these same systems except that connectedness is required on the accessibility relation $R$: $\mathbf{BNc}_{1,0}$, $\mathbf{BNc}_1$, $\mathbf{Bc}_{1,0}$, $\mathbf{Bc}_1$, $\mathbf{Nc}_{1,0}$, $\mathbf{Nc}_1$. All these systems are axiomatized and have soundness and completeness theorems (for two of the logics without the hereditary condition). This includes the one we have been discussing in this section, $\mathbf{BN}_{1,0}$. We will not list all the axiomatizations, but will describe just the one for $\mathbf{BN}_{1,0}$—to give the flavor.

The notion of "elliptical implication" features strongly in these axiomatizations. This was a notion first introduced by Robert Meyer, who called it *enthymematic implication*. It is defined using a logical constant $\boldsymbol{t}$, which Dunn (1966) had proved can be added conservatively to **R** with the axiom $A \leftrightarrow (\boldsymbol{t} \rightarrow A)$. So $\boldsymbol{t}$ can be thought of as the conjunction of all true sentences.

Dunn (2000) uses *elliptical implication*, $A \Rightarrow B$ to abbreviate $(A \wedge \boldsymbol{t}) \rightarrow ((A \wedge \boldsymbol{t}) \rightarrow B)$, and of course $A \Leftrightarrow B$ abbreviates $(A \Rightarrow B) \wedge (B \Rightarrow A)$. Then we can state two axioms, the axiom "Elliptical Negated Implication" ${\sim}(A \rightarrow B) \Leftrightarrow A \wedge {\sim}B$ and the axiom "Elliptical Self-Distribution" $(A \Rightarrow (B \Rightarrow C)) \Rightarrow ((A \Rightarrow B) \Rightarrow (A \Rightarrow C))$. These are valid in the above semantics, and with the rule "Elliptical Modus Ponens" $A \Rightarrow B,\ A \diamondsuit B$ give soundness and completeness for the logic $\mathbf{BN}_{1,0}$ when added to the axioms of **RW** (i.e., **R** minus the Contraction axiom).

None of these 12 logics is a modal logic, because each contains the so-called "Demodalizer" theorem $A \rightarrow ((A \rightarrow A) \rightarrow A)$. This is directly provable from $(A \rightarrow A) \rightarrow (A \rightarrow A)$ and Permutation, the relevant instance of which is $((A \rightarrow A) \rightarrow (A \rightarrow A)) \rightarrow (A \rightarrow ((A \rightarrow A) \rightarrow A))$ with self-implication as antecedent. The meaning of the demodalizer is that if $A$ is a true sentence then $A$ is necessary, that is, $A \rightarrow \Box A$, where $\Box$ is defined as $\Box A := (A \rightarrow A) \rightarrow A$.

## 6.5 A Modal Version of the Kripke-Style Semantics for RM

Given that the title of this volume mentions *modal logic*, it is about time that we examined a modal logic based on the Kripke-style semantics for **RM**. So we turn to the system obtained by adding the Mingle Axiom to the system **E** of Entailment. It is important to emphasize that this is the system **EM** and not the system $\mathbf{E}\overrightarrow{\mathbf{M}}$ with the restricted form of the Mingle Axiom, $\overrightarrow{A} \rightarrow (\overrightarrow{A} \rightarrow \overrightarrow{A})$. We will say more about this later. The system has already been studied in Dunn (1976c). Proof-theoretically,

[11] Slaney et al. (1989) discuss this same logic and call it simply **BN**. See also Brady (1982).

it is obtained by replacing the axiom of Permutation with the axiom of Restricted Permutation in the system **RM**. Model-theoretically, it is obtained by dropping the Trivalent Hereditary Condition altogether, and also dropping the requirement that the accessibility relation $R$ be a chain. It is still required to be a partial-order. Simultaneously dropping the two requirements makes some sense, since we have already noted that requiring that $R$ be connected was used in the proof of the Trivalent Hereditary Condition. And it is sensible to drop the Trivalent Hereditary Condition anyway since it is not a natural requirement for a modal logic. As observed in Dunn (1976c), the following kind of hereditary condition is easily proved for implications.

**Lemma 6.5.1** (Positive Hereditary Condition for Implications) *If* $HRH'$ *then, if* $t \in v(A \to B, H)$ *then* $t \in v(A \to B, H')$.

Furthermore, this is the only hereditary condition that is needed in verifying the axioms, which we will leave mostly to the reader. To make things easier (and more readable), instead of checking Restricted Permutation, we will prove Restricted Assertion $\overrightarrow{A} \to ((\overrightarrow{A} \to B) \to B)$, which is equivalent to it in the context of the other axioms of **EM** (indeed, already in **E**).

***Proof*** (*Restricted Assertion*) **1.** Let us assume that $t \in v(\overrightarrow{A}, H)$. We have to show that $t \in v((\overrightarrow{A} \to B) \to B, H)$. For the latter, we assume that $HRH'$. Then, first we also assume $t \in v(\overrightarrow{A} \to B, H')$. Since $t \in v(\overrightarrow{A}, H)$, by the Positive Hereditary Condition for Implications, we get $t \in v(\overrightarrow{A}, H')$. But this together with $t \in v(\overrightarrow{A} \to B, H')$ gives us $t \in v(B, H')$, what we had to prove.

We must also show that $f \in v(B, H')$ implies $f \in v(\overrightarrow{A} \to B, H')$, while still assuming $t \in v(\overrightarrow{A}, H)$ and $HRH'$. Again, we get $t \in v(\overrightarrow{A}, H')$, and since $f \in v(B, H')$, by the counterexample case, $f \in v(\overrightarrow{A} \to B, H')$.

**2.** Finally, we must prove that if $f \in v((\overrightarrow{A} \to B) \to B, H)$, then $f \in v(\overrightarrow{A}, H)$. You are, probably, getting used to this by now, and remember that there are two ways in which $f \in v((\overrightarrow{A} \to B) \to B, H)$ might hold. Either (counterexample) $t \in v(\overrightarrow{A} \to B, H)$ while $f \in v(B, H)$, or (non-truth) $t \notin v((\overrightarrow{A} \to B) \to B, H)$.

For the counterexample case, if $t \in v(\overrightarrow{A} \to B, H)$, but $f \in v(B, H)$, then for all $H'$ such that $HRH'$, if $f \in v(B, H')$ then $f \in v(\overrightarrow{A}, H')$. Of course, $HRH$ and since $f \in v(B, H)$, $f \in v(\overrightarrow{A}, H)$ as desired.

For the non-truth case, if $t \notin v((\overrightarrow{A} \to B) \to B, H)$, then for some $H'$ such that $HRH'$, either (i) $t \in v(\overrightarrow{A} \to B, H')$ and $t \notin v(B, H')$, or (ii) $f \in v(B, H')$ and $f \notin v(\overrightarrow{A} \to B, H')$. But we already know by **1** that if $t \in v(\overrightarrow{A}, H)$ were true, then so would be $t \in v((\overrightarrow{A} \to B) \to B, H)$. Hence, $t \notin v(\overrightarrow{A}, H)$, and so by Trivalence, $f \in v(\overrightarrow{A}, H)$ as we had to show. □

We now turn to completeness. First, we need to adjust the definition of the canonical model in ways familiar from modal logic. Thus, for prime theories $T'$ and $T''$ we define $T' < T''$ iff for all $A \to B$, if $A \to B \in T'$, then $A \to B \in T''$. Where $T$ is a prime **EM** theory, we define the *canonical model structure determined by* $T$ to be

$\langle G, K, R\rangle$, where $G = T$, $K = \{\, T : T$ is a prime **EM** theory and $G < T \,\}$, and $R$ is $<$. The corresponding *canonical valuation* $v$ is then defined so that (i) $t \in v(p, T')$ iff $p \in T'$, and (ii) $f \in v(p, T')$ iff $\sim p \in T'$. Then the *canonical model* is $\langle G, K, R, v\rangle$.

It is obvious that a canonical model structure so defined is in fact an **EM** model structure. What has to be verified is that clauses (i)–(ii) hold for all compound sentences $A$, and thereby, give us a valuation. An induction on the complexity of $A$ proceeds as for Theorem 6.2.10, until we come to the case, where $A$ is of the form $B \to C$. The argument for (ii) proceeds exactly as for **RM**, and the only subtle point is that we must have $(*)$ as a theorem of **EM**.

$$(*) \qquad (B \to C) \to \big((\sim B \vee C) \to (B \to C)\big)$$

A hurdle is that the proof of $(*)$ (within **RM**) given in (Dunn 1970, T11) depends on T9, that is, $(B \to C) \to \big(C \to (B \to C)\big)$, which in turn relies on Permutation. The latter is, of course, not a theorem of **EM**.

However, there is another proof of T9 that does not use Permutation, making it a theorem of **EM**. The first line is a theorem of $\mathbf{E}_{\to}$, and is easily proved in Kripke's sequent system $LE_{\to}$ from Kripke (1959b) (which we describe in the next section).

| | |
|---|---|
| 1. $[C \to (C \to C)] \to [(B \to C) \to (C \to (B \to C))]$ | $\mathbf{E}_{\to}$ theorem |
| 2. $C \to (C \to C)$ | Mingle axiom |
| 3. $(B \to C) \to (C \to (B \to C))$ | 1, 2, modus ponens |

To prove part (i) of the case of an implicational formula, we first prove the following lemma, which is a kind of deduction theorem for **EM**.

**Lemma 6.5.2** *Let* $\overrightarrow{\Gamma}$ *be a set of sentences of* **EM**, *all of the form* $C \to D$. *Then, if both* $\overrightarrow{\Gamma}$, $A \vdash_{\mathbf{EM}} B$ *and* $\overrightarrow{\Gamma}$, $\sim B \vdash_{\mathbf{EM}} \sim A$, *then* $\overrightarrow{\Gamma} \vdash_{\mathbf{EM}} A \to B$.

***Proof*** We note that the following enthymematic form of the deduction theorem is routinely provable for $\mathbf{EM}^{\boldsymbol{t}}$ (with $A \Rightarrow B$ defined as $(A \wedge \boldsymbol{t}) \to B$). The logical constant $\boldsymbol{t}$, intensional truth can be conservatively added to **EM** too—cf. Anderson and Belnap, 1975, pp. 342–343).

$$\text{If } \overrightarrow{\Gamma},\ A \vdash_{\mathbf{EM}} B,\ \text{ then } \overrightarrow{\Gamma} \vdash_{\mathbf{EM}} A \Rightarrow B.$$

Now, the claim follows from this enthymematic form by the same steps as in a similar proof in Dunn (1976a, p. 168). □

Finally, the following analog of Lemma 4 of Dunn (1976a) obviously suffices to establish (i) in the case of $B \to C$ in the induction.

**Lemma 6.5.3** $\vdash_T A \to B$ *iff for all prime* **EM** *theories* $T'$ *such that* $T < T'$, *if* $\vdash_{T'} A$, *then* $\vdash_{T'} B$, *and if* $\vdash_{T'} \sim B$ *then* $\vdash_{T'} \sim A$.

***Proof*** The left-to-right direction is immediate. For the right-to-left direction, suppose $\nvdash_T A \to B$. Set $\overrightarrow{T} = \{\, C \to D : \vdash_T C \to D \,\}$. Clearly $\overrightarrow{T} \nvdash_{\mathbf{EM}} A \to B$. But

then by Lemma 6.5.2, either $\overrightarrow{T},\ A \nvdash_{\mathbf{EM}} B$ or $\overrightarrow{T},\ {\sim}B \nvdash_{\mathbf{EM}} {\sim}A$. From here, the proof proceeds via a standard argument showing that either $\overrightarrow{T} \cup \{ A \}$ can be extended to a prime **EM** theory not containing $B$, or $\overrightarrow{T} \cup \{ {\sim}B \}$ can be extended to a prime theory while excluding ${\sim}A$. □

Two natural questions arise from the semantics for the system **EM**. The *first* is what happens if we require the relation $R$ to be a chain, i.e., what happens if we require that $R$ is connected (for all $H$, $H'$, $HRH'$ or $H'RH$)? We could entertain the thought that the well-known Chain Axiom from **RM** $(A \to C) \vee (C \to A)$, or perhaps, a twice restricted version $((A \to B) \to (C \to D)) \vee ((C \to D) \to (A \to B))$, which we label as (#), would work. Alas, neither formula is at least $t$ at $G$ in connected **EM** model structures as mentioned in Dunn (1976c, p. 66). Since **EM** is a *modal logic*, we might appropriate an axiom that is valid in a normal modal logic when $R$ is linear namely, $\Box(\Box A \supset C) \vee \Box(\Box C \supset A)$. We could contemplate replacing $\Box(\ \supset\ )$ with $\to$, since the latter has a modal flavor, and dropping the other $\Box$'s in lieu of $A$ and $C$ being implications; then we obtain (#). But we may recall that in **E** necessity is definable as $\Box A = (A \to A) \to A$. If we combine the two views on necessity, then we get to the formula $(((A \to A) \to A) \to C) \vee (((C \to C) \to C) \to A)$. We should also remember that $<$ is defined with reference to entailments; hence, we should take $A$ and $C$ to be entailments. Finally, we arrive at the following formula.

$$(\#\#)\quad \big((((A \to B) \to (A \to B)) \to (A \to B)) \to (C \to D)\big) \vee \big((((C \to D) \to (C \to D)) \to (C \to D)) \to (A \to B)\big).$$

The proof that each of (#) and (##) is sufficient to guarantee that $<$ is connected hints at the reason why we consider disjunctive formulas to start with.

***Proof*** (*of Connectedness*) We show that if $T < T_1$ and $T < T_2$ hold, then $T_2 < T_1$ or $T_1 < T_2$ also holds on the canonical model structure. Let us assume that neither $T_1 < T_2$ nor $T_2 < T_1$, which means that there are $A \to B \in T_1$ but $A \to B \notin T_2$, and $C \to D \in T_2$ but $C \to D \notin T_1$. Since $T$ is a prime theory, $(((A \to B) \to (A \to B)) \to (A \to B)) \to (C \to D) \in T$ or $(((C \to D) \to (C \to D)) \to (C \to D)) \to (A \to B) \in T$. There is a symmetry between the two cases, hence, we work out the details of the first. By assumption, $T < T_1$, hence, $(((A \to B) \to (A \to B)) \to (A \to B)) \to (C \to D) \in T_1$. However, the theorem $(A \to B) \to (((A \to B) \to (A \to B)) \to (A \to B))$ is an instance of Restricted Assertion, which is a theorem of **EM**. Then it is in $T_1$, and so $((A \to B) \to (A \to B)) \to (A \to B) \in T_1$, and finally $C \to D \in T_1$. This contradicts our initial assumption. (The steps using (#) are similar, even simpler.) □

Unfortunately, while both formulas would guarantee connectedness for $<$, neither is valid on connected **EM** models. We sketch some of the steps that would go into the verification of (##), then we formulate a conjecture.

Let us label $A \to B$ as $\varphi$ and $C \to D$ as $\psi$. Then we would like to show that $t \in v([((\varphi \to \varphi) \to \varphi) \to \psi] \vee [((\psi \to \psi) \to \psi) \to \varphi], G)$. Toward arriving at a

contradiction, let us assume $t \notin v([((\varphi \to \varphi) \to \varphi) \to \psi] \vee [((\psi \to \psi) \to \psi) \to \varphi], G)$, that is, $t \notin v([((\varphi \to \varphi) \to \varphi) \to \psi], G)$ and $t \notin v([((\psi \to \psi) \to \psi) \to \varphi], G)$. Then there is an $H$ such that $GRH$ and either $t \in v((\varphi \to \varphi) \to \varphi, H)$ and $t \notin v(\psi, H)$ or $f \in v(\psi, H)$ and $f \notin v((\varphi \to \varphi) \to \varphi, H)$. From the other conjunct, we get that there is an $H'$ such that $GRH'$ and either $t \in v((\psi \to \psi) \to \psi, H')$ and $t \notin v(\varphi, H')$ or $f \in v(\varphi, H')$ and $f \notin v((\psi \to \psi) \to \psi, H')$. By connectedness, either $HRH'$ or $H'RH$. Let's work with the first; the second is similar. There are four cases to consider by pairing the disjuncts of the conjuncts. We quickly sum up what we have. (i) $t \notin v(\psi, H)$, hence, $f \in v(\psi, H)$, and $t \in v(\varphi, H)$ as well as $t \notin v(\varphi, H')$, hence, $f \in v(\varphi, H')$, and $t \in v(\psi, H')$. (ii) Same facts in $H$ and $f \in v(\varphi, H')$, $f \notin v(\psi, H')$, $t \in v(\psi, H')$. (iii) $f \in v(\psi, H)$, $f \notin v(\varphi, H)$, hence, $t \in v(\varphi, H)$ as well as $f \in v(\varphi, H')$, $f \notin v(\psi, H')$ and $t \in v(\psi, H')$. And (iv) is same for $H$ as (iii) and $t \notin v(\varphi, H')$, $f \in v(\varphi, H')$ and $t \in v(\psi, H')$.

We remember that we have trivalence, but we only have the positive hereditary condition for implications by Lemma 6.5.1. Since $\varphi$ and $\psi$ are implicational formulas, we have a contradiction in cases (i) and (iv). In the other cases, we only have $f \notin v(\psi, H')$ to play with, because $t \in v(\varphi, H')$ and $f \in v(\varphi, H')$ is not a contradiction. We could contemplate extending Lemma 6.5.1, but a quick look at $(f\to)$ reveals that the falsity of an implicational formula at a state may not say anything about formulas at other states. (If we were to require the hereditary condition to hold for all formulas, then the proof would go through, but that would also mean that we are dealing with the semantics for **RM**.) A similar analysis of why (#) is not valid leads us to *conjecture* that there is no formula in the language of **EM** that would be valid on connected **EM** model structures and would force connectedness of the canonical model structure.

The *second* question is how to give a semantics for $\mathbf{E}\overrightarrow{\mathbf{M}}$. This was an open problem in Dunn (1976c). A partial answer to this question is given in the paper Kashima (2000), where an implicational **E**-mingle system is introduced. It is denoted by $\mathbf{EM}^{\#}$ in its axiomatic formulation, which adds to $\mathbf{E}_{\to}$ the Restricted Prefixed Mingle $(\overrightarrow{A} \to B) \to (\overrightarrow{A} \to (B \to B))$. A sequent calculus formulation is also defined and it is labeled as $\mathcal{G}\mathbf{EM}^{\#}$. The characteristic rule (beyond $\mathbf{E}_{\to}$) is

$$\frac{\Gamma, \overrightarrow{A} \vdash B}{\Gamma, \overrightarrow{A}, B \vdash B}\ \mathrm{m}^{\#}.$$

**Lemma 6.5.4** *$\mathcal{G}\mathbf{EM}^{\#}$ has the same set of theorems as $LE\overrightarrow{M}_{\to}$.*

***Proof*** We have not yet introduced $LE\overrightarrow{M}_{\to}$, but we will do that in the next section, in Definition 6.6.2. Restricted Prefixed Mingle is a theorem of $LE\overrightarrow{M}_{\to}$, and Restricted Mingle is a theorem of $\mathcal{G}\mathbf{EM}^{\#}$. We give the two easy derivations side by side.

$$\vdash\!\!\rightarrow,\ 2\times\ \frac{\mathrm{m}^{\#}\ \dfrac{\overrightarrow{A} \vdash \overrightarrow{A}}{\overrightarrow{A}, \overrightarrow{A} \vdash \overrightarrow{A}}}{\vdash \overrightarrow{A} \rightarrow (\overrightarrow{A} \rightarrow \overrightarrow{A})} \qquad \frac{\dfrac{\overrightarrow{A} \vdash \overrightarrow{A} \qquad B, B \vdash B}{\overrightarrow{A} \rightarrow B, \overrightarrow{A}, B \vdash B}\rightarrow\!\vdash}{\vdash (\overrightarrow{A} \rightarrow B) \rightarrow (\overrightarrow{A} \rightarrow (B \rightarrow B))}\ \vdash\!\!\rightarrow^{\varnothing},\ 3\times$$

□

A semantics for $\mathbf{EM}^{\#}$ is defined in the same paper, and it is an extension of the semi-lattice semantics for $\mathbf{E}_{\rightarrow}$ (see Urquhart (1972)). This is a semantics for the implicational fragment of $\mathbf{E}\overrightarrow{\mathbf{M}}_{\rightarrow}$, but it is not a variation on the binary relational semantics of **RM** or **EM** described above.

We leave further investigations of these questions for another occasion, and now, we turn to another problem that we solve.

## 6.6 The Intersection Problem

Motivations for the logic of entailment **E** include *necessity* and *relevance*—as both the title and the content of Anderson and Belnap (1975) and Anderson et al. (1992) reveal. However, necessity is not a primitive connective in the language of **E**. It is possible to define "necessary $A$" as $\Box A = \boldsymbol{t} \rightarrow A$ or from $\rightarrow$ alone as $\Box A = (A \rightarrow A) \rightarrow A$. This move leads to a concept of modality but it does not *explain* the modal character of $\rightarrow$. It is well known that $\mathbf{E}_{\rightarrow}$, the implicational fragment of **E** coincides with $\mathbf{R}^{\Box}_{\rightarrow}$, the modal implicational fragment of $\mathbf{R}^{\Box}$. This was proved in Meyer (1966). It is also well known that this translational characterization of the logic of entailment does not extend to the whole **E**. The latter was proved in Maksimova (1973).[12]

An explanation for **E**'s modal and relevant qualities could be given if **E** would turn out to be the intersection of a *modal logic* and of a *relevance logic*, for example, of **R**. As we mentioned in Sect. 6.1, a minglish extension of **E** has been suggested as the candidate to replace the modal logic **S4**.

At long last, after 56 years, we have a counterexample to the conjecture that $\mathbf{E}_{\rightarrow} = \mathbf{R}_{\rightarrow} \cap \mathbf{EM}_{\rightarrow}$. And now we unveil the counterexample (CE): $A \rightarrow [(A \rightarrow (A \rightarrow B)) \rightarrow (A \rightarrow B)]$. You will notice that this counterexample is the same as Kripke's to the conjecture that $\mathbf{E}_{\rightarrow} = \mathbf{R}_{\rightarrow} \cap \mathbf{S4}_{\rightarrow}$. Perhaps, this is why it was so difficult to see it as a counterexample to $\mathbf{E}_{\rightarrow} = \mathbf{R}_{\rightarrow} \cap \mathbf{EM}_{\rightarrow}$. It was hiding in plain sight as the saying goes. We did not just try Kripke's counterexample to see if it would work; we came up with "our own" counterexample, and then saw that it was the same as Kripke's.

It is easy to see that (CE) is a theorem of $\mathbf{R}_{\rightarrow}$. Permutation is often used as an axiom of $\mathbf{R}_{\rightarrow}$ (though only Restricted Permutation holds in $\mathbf{E}_{\rightarrow}$). It follows by Permutation, from an instance of the Identity axiom $(A \rightarrow (A \rightarrow B)) \rightarrow (A \rightarrow (A \rightarrow B))$ that

[12] Further connections between **E** and **S4** are described in Meyer (1970).

$A \to [(A \to (A \to B)) \to (A \to B)]$ is a theorem of $\mathbf{R}_{\to}$. But we have to show that it is a theorem of $\mathbf{EM}_{\to}$. Our proof will depend upon what Anderson and Belnap (1975, p. 26) call Replacement of the Middle: $(Y' \to Y) \to [X \to (Y \to Z)] \to [X \to (Y' \to Z)]$. Since this is a theorem of $\mathbf{E}_{\to}$ it is also a theorem of $\mathbf{EM}_{\to}$. (A proof can be easily constructed for this formula, because it is the principal type schema of the combinator $\mathsf{BBB}'$, which gives a recipe for a proof.)

1. $(A \to (A \to B)) \to (A \to B)$ Contraction
2. $(A \to A) \to ((A \to B) \to (A \to B))$ Suffixing
3. $A \to (A \to A)$ Mingle
4. $A \to ((A \to B) \to (A \to B))$ 3, 2 Transitivity
5. $A \to [(A \to (A \to B)) \to (A \to B)]$ Replacement of the Middle with

$$A \to (A \to B)/Y',\ A \to B/Y,\ A/X,\ A \to B/Z;\ \text{detachments}1, 4.$$

So (CE) is a theorem of both $\mathbf{R}_{\to}$ and $\mathbf{EM}_{\to}$, and hence an element of their intersection, but (CE) is not a theorem of $\mathbf{E}_{\to}$. The reason is (again) that its antecedent can be replaced by any sentence, including the propositional variable $p$. $\mathbf{E}_{\to}$ has the Ackermann property, which says that no theorem of $\mathbf{E}_{\to}$ is of the form in which a variable implies an implication. This was proved by Ackermann (cf. Anderson and Belnap, 1975, Sect. 5).

Now what happens if we consider the system $\mathbf{E}\overrightarrow{\mathbf{M}}$? Clearly, the proof above does not go through, because we only have restricted mingle $\overrightarrow{A} \to (\overrightarrow{A} \to \overrightarrow{A})$. But, if we require in the above proof that $A$ be an implication $\overrightarrow{A} := C \to D$, then it appears that we might have a counterexample to $\mathbf{E}_{\to} = \mathbf{R}_{\to} \cap \mathbf{E}\overrightarrow{\mathbf{M}}_{\to}$ too. However, we do not have the same easy proof to show that $\overrightarrow{A} \to [(\overrightarrow{A} \to (\overrightarrow{A} \to B)) \to (\overrightarrow{A} \to B)]$ is not a theorem of $\mathbf{E}_{\to}$. Indeed, we do have a proof that the formula is a theorem of $LE_{\to}$, that is, of $\mathbf{E}_{\to}$. The following is a proof in Kripke's original calculus $LE_{\to}$ from Kripke (1959b), which we describe in Definition 6.6.1. (And the proof should already look like a correct one—at least in $LK$.)

$$\dfrac{\dfrac{\dfrac{\dfrac{C \to D \vdash C \to D \qquad (C \to D) \to B \vdash (C \to D) \to B}{(C \to D) \to ((C \to D) \to B), C \to D \vdash (C \to D) \to B}}{C \to D \vdash ((C \to D) \to ((C \to D) \to B)) \to ((C \to D) \to B)}}{\vdash (C \to D) \to \big(((C \to D) \to ((C \to D) \to B)) \to ((C \to D) \to B)\big)}$$

Anderson and Belnap motivated the precursor of the $\mathbf{E}_{\to} = \mathbf{R}_{\to} \cap \mathbf{E}\overrightarrow{\mathbf{M}}_{\to}$ conjecture by the observation that combining two kinds of restrictions on a natural deduction system yields $\mathbf{FE}_{\to}$. We might ask whether it is possible to view $LE_{\to}$ (or $LE^{t}_{\to}$) as a sequent calculus that satisfies a pair of restrictions related to $\mathbf{R}_{\to}$ and $\mathbf{E}\overrightarrow{\mathbf{M}}_{\to}$.[13]

[13] Nowadays, there are several sequent calculi for $\mathbf{E}_{\to}$ and for some of its conservative extensions. Belnap and Wallace (1965) add negation; Anderson and Belnap (1975) present a merge system;

Belnap seemed to have favored the use of indices on formulas, which is not completely unlike how indices are used in $\mathbf{FE}_{\rightarrow}$ (see Anderson and Belnap (1975, Sect. 5)). In a letter to Belnap (dated June 13th, 1959), Kripke outlined his sequent calculus for $\mathbf{E}_{\rightarrow}$, which he described (in words) in Kripke (1959b).[14] Notation and terminology used by Kripke suggest that he was familiar with Curry (1950), and he saw the path to $\mathbf{E}_{\rightarrow}$ from $\mathbf{R}_{\rightarrow}$ through a side condition on the right introduction rule for $\rightarrow$, which he denoted by Pr. Curry used a side condition in the rule Pl in his definition of a sequent calculus for positive intuitionistic logic $LA(\mathfrak{S})$ in Curry (1950, II. 4). (Later, he also added a side condition to P$*$ in $LA_m$ in Curry, 1963, 5C4.) Freeing the calculus from indices attached to formulas while avoiding the merge operation on sequences of formulas in the left introduction rule for implication is the first remarkable innovation introduced by Kripke toward a decidability proof.

**Definition 6.6.1** (Kripke, 1959b) Sequents are ordered pairs comprising a sequence of formulas and a formula. We write $\Gamma \vdash A$ instead of $\langle \Gamma, A \rangle$—as usual. $\Gamma$, the left-hand side of the sequent may be empty, but the right-hand side of a sequent is never empty.

The sequent calculus $LE_{\rightarrow}$ comprises the axiom (id) and the following rules.

$$A \vdash A \ \text{id}$$

$$\frac{\Gamma \vdash A \quad B, \Theta \vdash C}{A \rightarrow B, \Gamma, \Theta \vdash C} \rightarrow\vdash \qquad \frac{\overrightarrow{\Gamma}, A \vdash B}{\overrightarrow{\Gamma} \vdash A \rightarrow B} \vdash\rightarrow^{\oslash}$$

$$\frac{\Gamma, A, B, \Theta \vdash C}{\Gamma, B, A, \Theta \vdash C} C\vdash \qquad \frac{\Gamma, A, A \vdash B}{\Gamma, A \vdash B} W\vdash$$

The labels for the structural rules allude to combinators, that is, $(C \vdash)$ stands for the permutation rule and $(W \vdash)$ stands for the contraction rule. (A similar notational convention was introduced by Curry.) The $\overrightarrow{\Gamma}$ notation means that every element of $\Gamma$ is an implicational formula (i.e., not an atomic formula). The superscript $^{\oslash}$ indicates that the rule is a restricted version of the usual right introduction rule for $\rightarrow$, namely, it requires $\overrightarrow{\Gamma}$ rather than $\Gamma$. It is a crucial insight of Kripke into $\mathbf{E}_{\rightarrow}$ that permutation may be included as a rule, and instead of limiting permutation, a constraint on Pr can produce restricted versions of some $\mathbf{R}$ theorems (which are theorems of $\mathbf{E}$). Since permutation is fully available, we have not been too fussy about the location of the principal formulas and their subalterns in the implication and contraction rules. It is easy to see that Restricted Permutation is provable, because moving implicational formulas toward left is in harmony with $(\vdash\rightarrow^{\oslash})$.

The sequent calculus "dual" of contraction is close to weakening, but it's not exactly weakening, it is *expansion*. Kripke in a letter to Belnap (dated June 13th, 1959), considered the above sequent calculus with the rule of expansion added, and stated that ET, that is, the elimination theorem, in Curry's terminology, holds.

---

Bimbó (2009) and Bimbó (2015) add $\boldsymbol{t}$ (in two ways) as well as $\wedge$ and $\vee$ to formalize $\mathbf{E}^{\boldsymbol{t}}_{+}$. However, we stress that the *first* sequent calculus for the implicational fragment was defined by Kripke.

[14] A copy of the letter is preserved in the Kripke Archives.

The two extensions of $LE_{\to}$ that we consider are $LE\overrightarrow{M}_{\to}$ and $LR_{\to}$.

**Definition 6.6.2** $LE\overrightarrow{M}_{\to}$ is an extension of $LE_{\to}$ in which the axiom (id) is replaced by (id$^n$), that is, $A^n \vdash A$, the superscript indicating $n \in \mathbb{N}^+$ occurrences of $A$. $LR_{\to}$ is an extension of $LE_{\to}$ in which the $(\vdash\to)$ rule has no restrictions, that is, $\overrightarrow{\Gamma}$ is replaced by $\Gamma$ in both the upper and the lower sequents.

**Theorem 6.6.3** (Kripke, 1959b) *The cut rule is admissible in $LE\overrightarrow{M}_{\to}$ and in $LR_{\to}$.*

The cut theorem for $LR_{\to}$ was stated in Kripke (1959b). Although Kripke in the letter mentioned above stated the cut theorem, he used the expansion rule rather than an axiom. Avron (1987) considers a sequent calculus $GRM^{\sim}_{\to}$ that uses a similar axiom. The additional connectives ($\sim$, $+$ and $\circ$) and the multiple right-hand sides do not interfere with the proof of the cut theorem, nor does the restriction on $(\vdash\to^{\oslash})$. We will not spell out the details of the proof here, but refer to the proofs mentioned.

**Remark 1** We note some crucial points concerning the proof though.

*First* of all, the multiple cut is known to be equivalent to the single cut where there are rules such as the contraction and permutation rules (cf. Bimbó (2007a)). The use of multiple cut is unproblematic where the only connective is implication. $C^n$ is one or more (but finitely many) occurrences of $C$.

$$\frac{\Gamma \vdash C \qquad C^n, \Theta \vdash A}{\Gamma, \Theta \vdash A}\ \text{multi cut}$$

*Second*, it may be assumed that the transformations of a proof are carried out relying on a strategy, in particular, the left rank is reduced whenever possible. *Third*, the proof can proceed by double induction with the degree of a formula defined as usual, and the rank defined with reference to parametric ancestors (not simply occurrences) of a formula. *Fourth*, we note that a cut with a pair of axioms results in an axiom; however, it may not be identical to either premise of the cut. (The following illustrates the case with $n, m \in \mathbb{N}^+$.)

$$\frac{A^n \vdash A \qquad A^m \vdash A}{A^{n+m-1} \vdash A}$$

*Fifth*, if the right premise is an instance of the axiom and the left premise is introduced by the $(\vdash\to)$ rule, then we have to modify the proof (unlike in the usual case). On the left is the given proof, and on the right is the modified proof. The new cut is justified by a reduction in the degree of the cut formula.

$$\frac{\dfrac{\Gamma, A \vdash B}{\Gamma \vdash A \to B} \qquad A \to B, A \to B \vdash A \to B}{\Gamma, A \to B \vdash A \to B}$$

$$\dfrac{\dfrac{A \vdash A \qquad \dfrac{\Gamma, A \vdash B \qquad B, B \vdash B}{\Gamma, A, B \vdash B}}{\Gamma, A \to B, A, A \vdash B}}{\dfrac{\Gamma, A \to B, A \vdash B}{\Gamma, A \to B \vdash A \to B}}$$

**Lemma 6.6.4** *The axiom* $(\mathrm{id}^n)$ *is not derivable in* $LR_{\to}$ *and the rule* $\vdash\!\to$ *is not admissible in* $LE\overrightarrow{M}_{\to}$.

***Proof*** For the first claim, we note that $p \to (p \to p)$ is not a theorem of $LR_{\to}$. We will provide two semantic arguments.[15] The first appeals to the matrix $M_0$, which is well-known in the relevance logic literature. (See, for example, Anderson and Belnap 1975, Sect. 26); we will make more use of this matrix later in this section.) If $p$ is given the value $+2$, then the self-implication gets the value $+0$, and the whole formula is $-3$. Alternatively, we may use the Meyer–Routley semantics for $R^{\circ}_{\to}$. Having $R\alpha\beta\gamma$ with $\alpha, \beta \in v(p)$ and $\gamma \notin v(p)$ yields that some $\iota \in I$ does not satisfy the formula. The stipulations do not contradict anything in the definition of the semantics, and the four situations can be closed under the defining conditions to generate a model.

For the second claim, we note that $(p \to (q \to r)) \to (q \to (p \to r))$ is not a theorem of $LE\overrightarrow{M}_{\to}$. (We assume that $p, q$ and $r$ are distinct propositional variables.) This may be shown using a bottom-up proof search. It is sufficient to consider cut-free proofs, which severely limits the number of possible steps. □

**Theorem 6.6.5** *If a proof is a proof in* $LR_{\to}$ *as well as in* $LE\overrightarrow{M}_{\to}$, *then it is a proof in* $LE_{\to}$.

***Proof*** The claim is practically obvious, because we defined $LR_{\to}$ and $LE\overrightarrow{M}_{\to}$ to be extensions of $LE_{\to}$. Moreover, the extensions are "orthogonal," that is, neither extension is an extension of the other extension. □

This theorem characterizes $LE_{\to}$ as an intersection of two of its extensions. However, the theorem concerns *proofs* not formulas, hence, it's facile. Another trite theorem says that in an axiomatic calculus (assuming the usual lax notion of a proof) any provable formula has infinitely many distinct proofs. Sequent calculi, especially, sequent calculi formalizing a relevance logic may preclude the insertion of arbitrary formulas to generate infinitely many distinct proofs for a formula. (Indeed, in $LR_{\to}$ and $LE_{\to}$ the number of cut-free irredundant proofs for a theorem is finite; still it can be more than one.) However, the above theorem does not generalize to formulas if there are at least two proofs for at least one formula.

Meyer (1976) noted that if a formula $A$ is necessary (in the sense of Anderson and Belnap), then in $\mathbf{E}_{\to}$ it implies an arbitrary $B$ (and itself too) on the condition that $A$ and $B$ coimply each other. In fact, this already happens in the logic of ticket entailment $\mathbf{T}_{\to}$, which is a "sublogic" of $\mathbf{E}_{\to}$. The following two proofs are in $LT^{t}_{\to}$, introduced in Bimbó (2007b), which contains more information about $\mathbf{T}$.[16]

[15] We could construct a proof-search tree, etc., but either way we have to rely on a lot of known facts here, which we cannot reprove in this paper. We refer to Dunn (1986) and Anderson et al. (1992) for more information on $\mathbf{R}$ and $\mathbf{E}$.

[16] $\mathbf{T}_{\to}$ is also decidable; cf. Bimbó and Dunn (2012) and Bimbó and Dunn (2013).

$$\dfrac{\dfrac{\dfrac{\dfrac{\dfrac{\dfrac{\dfrac{\dfrac{}{A \vdash A} \quad \dfrac{\dfrac{}{B \vdash B} \quad \dfrac{}{A \vdash A}}{B \to A; B \vdash A}}{B \to A; (A \to B; A) \vdash A}}{(A \to B; B \to A); A \vdash A}}{A \to B; B \to A \vdash A \to A} \quad \dfrac{}{A \vdash A}}{(A \to A) \to A; (A \to B; B \to A) \vdash A}}{((A \to A) \to A; A \to B); B \to A \vdash A}}{(A \to A) \to A \vdash (A \to B) \to ((B \to A) \to A)}$$

$$\dfrac{\dfrac{\dfrac{\dfrac{\dfrac{\dfrac{\dfrac{\dfrac{\dfrac{}{B \vdash B} \quad \dfrac{\dfrac{}{A \vdash A} \quad \dfrac{}{B \vdash B}}{A \to B; A \vdash B}}{A \to B; (B \to A; B) \vdash B}}{(A \to B; B \to A); B \vdash B}}{A \to B; B \to A \vdash B \to B} \quad \dfrac{}{B \vdash B}}{(B \to B) \to B; (A \to B; B \to A) \vdash B}}{((B \to B) \to B; A \to B); B \to A \vdash B} \quad \dfrac{}{A \vdash A}}{B \to A; ((B \to B) \to B; A \to B; B \to A) \vdash A}}{(((B \to B) \to B; A \to B); B \to A); B \to A \vdash A}}{((B \to B) \to B; A \to B); B \to A \vdash A}}{(B \to B) \to B \vdash (A \to B) \to ((B \to A) \to A)}$$

The formulas that we retained on the left-hand side of the turnstile express $\Box A$ and $\Box B$, respectively. The formula on the right-hand side of the turnstile is the same one; that is, if $A$ and $B$ are both necessary, then $(A \to B) \to ((B \to A) \to A)$ has (at least) two different proofs.

A natural source of necessary formulas is the pool of theorems. Since $A \to A$ is a theorem of all relevance logics, from $\Box A$, $A$ follows. And if $A$ is a theorem, then $A \to A$ (or $\boldsymbol{t}$) should imply it, because $A \to A$ is relevant to $A$.[17] This informal reasoning is reflected by the following proof in $LE^{\boldsymbol{t}}_{\to}$.

$$\dfrac{\dfrac{\dfrac{\dfrac{}{\boldsymbol{t} \vdash A} \quad \dfrac{}{A \vdash A}}{A \to A; \boldsymbol{t} \vdash A}}{A \to A \vdash A}}{\boldsymbol{t}; A \to A \vdash A}}{\boldsymbol{t} \vdash (A \to A) \to A}$$

[17] The idea of simulating $\boldsymbol{t}$ by conjunctions of self-implications among the propositional variables of a theorem goes back to Ackermann (1956).

From the point of view of $\mathbf{E}\overrightarrow{\mathbf{M}}_{\rightarrow}$, $\overrightarrow{M}$ is necessary, and from the point of view of $\mathbf{R}_{\rightarrow}$, $A \rightarrow ((A \rightarrow B) \rightarrow B)$ is necessary. From the point of view of $\mathbf{E}_{\rightarrow}$, these formulas are characteristic for those respective calculi. We will denote a particular instance of $\overrightarrow{M}$ as $\varphi$, and that of $A \rightarrow ((A \rightarrow B) \rightarrow B)$ as $\psi$, and we assume that $\varphi$ and $\psi$ share no variables. For the sake of concreteness, $\varphi$ is $(p \rightarrow q) \rightarrow ((p \rightarrow q) \rightarrow (p \rightarrow q))$ and $\psi$ is $r \rightarrow ((r \rightarrow s) \rightarrow s)$, where $p, q, r$ and $s$ are distinct. ($[\![\ ]\!]$ gives the value of a formula.)

**Lemma 6.6.6** *$M_0 \vDash \psi$ and $M_0 \nvDash \varphi$. Specifically, if the values for the propositional variables $[\![p]\!] = +0$ and $[\![q]\!] = -0$, then $[\![\varphi]\!] = -3$.*

The proof is an easy calculation on $M_0$, and we leave it to the reader.

Some more calculation on $M_0$ also reveals that $(A \rightarrow B) \rightarrow ((B \rightarrow A) \rightarrow A)$ can only be refuted with both $[\![A]\!], [\![B]\!] \in \{-0, -1, -2, -3\}$ and $[\![A]\!] = [\![B]\!]$. In other words, $M_0$ cannot refute the instance of this formula where $A$ and $B$ are $\varphi$ and $\psi$ (in either order).

Now, *modality* comes to our rescue. Dunn (1966) algebraized $\mathbf{R}^{t\Box}$ as a *De Morgan monoid with a closure operator* $\mathsf{C}$. The latter operation satisfies (c1)–(c5). ($a, b \in A$, the carrier set of the algebra, and $e$ is the monoid identity; $\sim$, $\vee$ and $\rightarrow$ are the type-lifted operations obtained from the respective connectives.)

(c1) $a \leq \mathsf{C}\, a$
(c2) $\mathsf{C}\,\mathsf{C}\, a = \mathsf{C}\, a$
(c3) $\mathsf{C}(a \vee b) = \mathsf{C}\, a \vee \mathsf{C}\, b$
(c4) $\mathsf{C}(\sim e) = \sim e$
(c5) $\sim \mathsf{C} \sim (a \rightarrow b) \leq \sim \mathsf{C} \sim a \rightarrow \sim \mathsf{C} \sim b$.

**Lemma 6.6.7** *Let $\Box$ be defined on $M_0$ as follows. Then $M_0^{\Box}$ satisfies (c1)–(c5).*

| | $-3$ | $-2$ | $-1$ | $-0$ | $+0$ | $+1$ | $+2$ | $+3$ |
|---|---|---|---|---|---|---|---|---|
| $\Box$ | $-3$ | $-3$ | $-3$ | $-0$ | $+0$ | $+0$ | $+0$ | $+3$ |

Once again, we leave the easy (but a bit tedious) checking to the reader.

Now, we recall that $\mathbf{E}$ is a modal logic, and it was conjectured by Meyer (1966) that if $A \rightarrow_{\mathbf{E}} B$ is translated as $\Box(A \rightarrow_{\mathbf{R}^{\Box}} B)$, then $\mathbf{E}$ would emerge as the exact translational fragment of $\mathbf{R}^{\Box}$. This conjecture holds for the implicational fragment, as Meyer (1966) proved, and it fails for $\mathbf{E}$ (even for $\mathbf{E}_{+}$) as Maksimova (1973) proved. Luckily, for our purposes here, the implicational result suffices. The modal translations of $\varphi$ and $\psi$ will be denoted by $\varphi^{\Box}$ and $\psi^{\Box}$, respectively.

($\varphi^{\Box}$) $\Box(\Box(p \rightarrow q) \rightarrow \Box(\Box(p \rightarrow q) \rightarrow \Box(p \rightarrow q)))$
($\psi^{\Box}$) $\Box(r \rightarrow \Box(\Box(r \rightarrow s) \rightarrow s))$

**Lemma 6.6.8** (1) *$M_0^{\Box} \nvDash \varphi^{\Box}$ and* (2) *$M_0^{\Box} \nvDash \psi^{\Box}$.*

***Proof*** For (1), let $[\![p]\!] = +0$ and $[\![q]\!] = -0$. Then $[\![\varphi^{\Box}]\!] = -3$. For (2), there are several pairs of values that work. For instance, $[\![r]\!] = +2$ and $[\![s]\!] = +2$ yield $[\![\psi^{\Box}]\!] = -3$. (The other pairs are $\langle -2, +2\rangle$, $\langle -2, -2\rangle$, $\langle -1, -1\rangle$, $\langle -1, +1\rangle$ and $\langle +1, +1\rangle$, each producing the value $-3$.) We leave it to the reader to crunch the numbers. □

Let us now take stock of what we have so far. We have shown that—as we expected—neither $\varphi$ nor $\psi$ are theorems of $\mathbf{E}_{\to}$. However, they are theorems of $\mathbf{E}\overrightarrow{\mathbf{M}}_{\to}$ and $\mathbf{R}_{\to}$, respectively. Then $\Box\varphi$ and $\Box\psi$ are also theorems of those logics. We have proved that $\Box A$ and $\Box B$ imply $(A \to B) \to ((B \to A) \to A)$ in $\mathbf{T}_{\to}$, hence, in $\mathbf{E}_{\to}$ too. $\mathbf{E}\overrightarrow{\mathbf{M}}_{\to}$ and $\mathbf{R}_{\to}$ are extensions of $\mathbf{E}_{\to}$, hence,

$$(\dagger) \qquad (\psi \to \varphi) \to ((\varphi \to \psi) \to \psi)$$

is a theorem of both. It remains to show that (†) is not a theorem of $\mathbf{E}_{\to}$.

**Lemma 6.6.9** $M_0^{\Box} \not\models \Box(\Box(A \to B) \to \Box(\Box(B \to A) \to A))$.

***Proof*** The following pairs for $\langle [\![A]\!], [\![B]\!]\rangle$ prove the claim: $\langle -0, -0\rangle \mapsto -0$, and each of $\langle -1, -1\rangle$, $\langle -2, -2\rangle$ and $\langle -3, -3\rangle$ gives $-3$. □

We have found a range of values for a refutation, so we need just one more step.

**Theorem 6.6.10** $\mathbf{E}_{\to} \not\vdash (\psi \to \varphi) \to ((\varphi \to \psi) \to \psi)$.

***Proof*** We use once again the modally enhanced $M_0^{\Box}$ and the translation of our formula (†), which is

$$(\dagger^{\Box}) \qquad \Box(\Box(\psi^{\Box} \to \varphi^{\Box}) \to \Box(\Box(\varphi^{\Box} \to \psi^{\Box}) \to \psi^{\Box})).$$

By Lemma 6.6.9, we know that it is sufficient to choose values for the propositional variables that will yield $-3$ for both $\psi^{\Box}$ and $\varphi^{\Box}$. By Lemma 6.6.8, we know that $[\![p]\!] = +0$, $[\![q]\!] = -0$, $[\![r]\!] = +2$ and $[\![s]\!] = +2$ do the job.

This completes the proof that $\mathbf{E}_{\to} \neq \mathbf{R}_{\to} \cap \mathbf{E}\overrightarrow{\mathbf{M}}_{\to}$. □

## 6.7 Conclusions

In this paper, we explored logics that are related to implicational **E**, a *modal relevance logic* that was given an elegant *sequent calculus* formalization by Kripke, who also proved it *decidable*. Kripke (1959b) describes the sequent calculus and states the decidability of $\mathbf{E}_{\to}$. Kripke communicated further details of his proof to Nuel D. Belnap in a letter dated September 21st, 1959. Details of the proof—based on Kripke's letter—can be found in print in Anderson and Belnap (1975, Sect. 13), as well as in Dunn (1986, Sect. 3.6). Kripke's insights made earlier methods of decidability proofs more flexible and applicable to relevance and other logics. This is

detailed in Bimbó (2020), where the evolution of sequent calculus based decidability proofs is described. Our explorations of modal versions of **RM** led us to the solution of the original problem posed by McCall in connection to the intersection problem that Kripke solved (in October 1959) concerning whether the "S4ish" modality of $\mathbf{E}_{\rightarrow}$ emerges as the intersection of $\mathbf{S4}_{\rightarrow}$ and $\mathbf{R}_{\rightarrow}$.

*We present this article as an intellectual tribute to Saul Kripke in his 81st year.*

**Acknowledgements** We would like to thank the editors for inviting us to write an article for this volume and two anonymous referees for their useful comments.
We are grateful to Yale Weiss, who is the Assistant Director of the Kripke Center at the City University of New York, for providing us with copies of some of Saul Kripke's correspondence (kept in the Kripke Archive at CUNY) that pertains to his work on relevance logics.
We acknowledge financial support during the writing of this paper from the *Insight Grant* #435–2019–0331 awarded by the *Social Sciences and Humanities Research Council of Canada*.

## Appendix

In order to make this paper more self-contained, we list a series of formulas, which are often taken to be axioms for various logics. (From (A1) to (A18) we follow the numbering in Anderson et al., 1992, Sect. R2).)

| | | |
|---|---|---|
| (A1) | $A \rightarrow A$ | [pt(I)] |
| (E1) | $((A \rightarrow A) \rightarrow B) \rightarrow B$ | [pt(CII)] |
| (A2) | $(A \rightarrow B) \rightarrow ((B \rightarrow C) \rightarrow (A \rightarrow C))$ | [pt(B′)] |
| (A3) | $(A \rightarrow B) \rightarrow ((C \rightarrow A) \rightarrow (C \rightarrow B))$ | [pt(B)] |
| (R3) | $A \rightarrow ((A \rightarrow B) \rightarrow B)$ | [pt(CI)] |
| (A4) | $(A \rightarrow (A \rightarrow B)) \rightarrow (A \rightarrow B)$ | [pt(W)] |
| (A5) | $(A \wedge B) \rightarrow A$ | |
| (A6) | $(A \wedge B) \rightarrow B$ | |
| (A7) | $((A \rightarrow B) \wedge (A \rightarrow C)) \rightarrow (A \rightarrow (B \wedge C))$ | |
| (A8) | $A \rightarrow (A \vee B)$ | |
| (A9) | $B \rightarrow (A \vee B)$ | |
| (A10) | $((A \rightarrow B) \wedge (C \rightarrow B)) \rightarrow ((A \vee C) \rightarrow B)$ | |
| (A11) | $(A \wedge (B \vee C)) \rightarrow ((A \wedge B) \vee C)$ | |
| (A12) | $(A \rightarrow {\sim}A) \rightarrow {\sim}A$ | |
| (A13) | $(A \rightarrow {\sim}B) \rightarrow (B \rightarrow {\sim}A)$ | |
| (A14) | ${\sim}{\sim}A \rightarrow A$ | |
| (A16) | $[((A \rightarrow A) \rightarrow A) \wedge ((B \rightarrow B) \rightarrow B)] \rightarrow [((A \wedge B) \rightarrow (A \wedge B)) \rightarrow (A \wedge B)]$ | |
| (A18) | $(A \rightarrow B) \rightarrow ((A \rightarrow B) \rightarrow (A \rightarrow B))$ | |
| (M) | $A \rightarrow (A \rightarrow A)$ | |
| (E$t$) | $(t \rightarrow A) \rightarrow A$ | |

(MP) $A \to B,\ A \ \Diamond\ B$
(Ad) $A,\ B \ \Diamond\ A \wedge B$
($\boldsymbol{t}$) $\vdash A$ iff $\vdash \boldsymbol{t} \to A$

Taking all these formulas as axioms and adding the three rules, we obtain $\mathbf{RM}^{\boldsymbol{t}}$. Omitting (E$\boldsymbol{t}$) and ($\boldsymbol{t}$) gives **RM**. If we further drop (M) and (A18) we get **R** (and $\boldsymbol{t}$ may be added back by ($\boldsymbol{t}$), if desired). **E** results by excluding assertion (R3) from **R**. Intensional truth may be added to **E** by (E$\boldsymbol{t}$) and ($\boldsymbol{t}$). Restricted mingle (A18) and mingle (M) yield two distinct extensions of **E**. **T** is axiomatized by (A1)–(A14) with (MP) and (Ad); $\boldsymbol{t}$ can be added by ($\boldsymbol{t}$).

## References


Ackermann, W. (1956). Begründung einer strengen Implikation. *Journal of Symbolic Logic, 21*, 113–128.

Anderson, A. R., & Belnap, N. D. (1975). *Entailment: The logic of relevance and necessity* (Vol. I). Princeton, NJ: Princeton University Press.

Anderson, A. R., Belnap, N. D., & Dunn, J. M. (1992). *Entailment: The logic of relevance and necessity* (Vol. II). Princeton, NJ: Princeton University Press.

Avron, A. (1987). A constructive analysis of RM. *Journal of Symbolic Logic, 52*(4), 939–951.

Avron, A. (2016). Paraconsistent fuzzy logic preserving non-falsity. *Fuzzy Sets and Systems, 292*, 75–84.

Beall, J., Brady, R., Dunn, J. M., Hazen, A. P., Mares, E., Meyer, R. K., & Sylvan, R. (2012). On the ternary relation and conditionality. *Journal of Philosophical Logic, 41*, 595–612.

Belnap, N. D. (1959). *The formalization of entailment* (Unpublished doctoral dissertation). New Haven, CT: Yale University.

Belnap, N. D., & Wallace, J. R. (1965). A decision procedure for the system $E_{\overline{I}}$ of entailment with negation. *Zeitschrift für mathematische Logik und Grundlagen der Mathematik, 11*, 277–289.

Bimbó, K. (2007a). $LE^{t}_{\to}$, $LR^{\circ}_{\mathrel{\wedge\!\!\sim}}$, LK and cutfree proofs. *Journal of Philosophical Logic, 36*(5), 557–570.

Bimbó, K. (2007b). Relevance logics. In D. Jacquette (Ed.), *Philosophy of logic*. *Handbook of the philosophy of science* (Vol. 5, pp. 723–789). Amsterdam: Elsevier (North-Holland).

Bimbó, K. (2009). Dual gaggle semantics for entailment. *Notre Dame Journal of Formal Logic, 50*(1), 23–41.

Bimbó, K. (2015). *Proof theory: Sequent calculi and related formalisms*. Boca Raton, FL: CRC Press.

Bimbó, K. (2020). The development of decidability proofs based on sequent calculi. In A. Rezuş (Ed.), *Contemporary logic and computing*. *Landscapes in logic* (Vol. 1, pp. 5–37). London, UK: College Publications.

Bimbó, K., & Dunn, J. M. (2012). New consecution calculi for $R^{\mathbf{t}}_{\to}$. *Notre Dame Journal of Formal Logic, 53*(4), 491–509.

Bimbó, K., & Dunn, J. M. (2013). On the decidability of implicational ticket entailment. *Journal of Symbolic Logic, 78*(1), 214–236.

Bimbó, K., & Dunn, J. M. (2017). The emergence of set-theoretical semantics for relevance logics around 1970. In K. Bimbó & J. M. Dunn (Eds.), *Proceedings of the Third Workshop, May 16-17, 2016, Edmonton, Canada*. *IFCoLog Journal of Logics and Their Applications, 4*(3), 557–589. London, UK: College Publications.

Brady, R. T. (1982). Axiomatisation of RM3 and BN4. *Logique et Analyse, 25*, 9–32.

Curry, H. B. (1950). *A theory of formal deducibility*. *Notre Dame mathematical lectures* (No. 6). Notre Dame, IN: University of Notre Dame Press.
Curry, H. B. (1963). *Foundations of mathematical logic*. New York, NY: McGraw-Hill Book Company. (Dover, New York, NY, 1977.)
Dretske, F. (1981). *Knowledge and the flow of information*. Cambridge, MA: MIT Press.
Dunn, J. M. (1966). *The algebra of intensional logics* (Doctoral dissertation). University of Pittsburgh, Ann Arbor (UMI). (Published as Vol. 2 in the Logic PhDs series by College Publications, London (UK), 2019.)
Dunn, J. M. (1970). Algebraic completeness results for R-mingle and its extensions. *Journal of Symbolic Logic, 35*(1), 1–13.
Dunn, J. M. (1976a). A Kripke-style semantics for R-mingle using a binary accessibility relation. *Studia Logica, 35*, 163–172.
Dunn, J. M. (1976b). Review of 'Meyer, R. K., Two questions from Anderson and Belnap. *Reports on Mathematical Logic, 7*(1976), 71–86'. Mathematical Reviews MR0505276 (58 #21478).
Dunn, J. M. (1976c). A variation on the binary semantics for R-Mingle. *Relevance Logic Newsletter, 1*(2), 56–67.
Dunn, J. M. (1986). Relevance logic and entailment. In D. Gabbay & F. Guenthner (Eds.), *Handbook of philosophical logic* (1st ed., Vol. 3, pp. 117–224). Dordrecht: D. Reidel.
Dunn, J. M. (2000). Partiality and its dual. *Studia Logica, 65*, 5–40.
Dunn, J. M. (2013). Guide to the Floridi keys: Essay review of Luciano Floridi's The Philosophy of Information. *MetaScience, 22*, 93–98.
Dunn, J. M. (2021). R-Mingle is nice, and so is Arnon Avron. In O. Arieli & A. Zamansky (Eds.), *Arnon Avron on semantics and proof theory of non-classical logics*. *Outstanding Contribution to Logic* (Vol. 21, pp. 141–165). Switzerland: Springer Nature.
Dunn, J. M., & Restall, G. (2002). Relevance logic. In D. Gabbay & F. Guenthner (Eds.), *Handbook of philosophical logic* (2nd ed., Vol. 6, pp. 1–128). Amsterdam: Kluwer.
Fetzer, J. (2004). Information: Does it have to be true? *Minds and Machines, 14*(2), 223–229.
Floridi, L. (2011). *The philosophy of information*. Oxford (UK): Oxford University Press.
Grzegorczyk, A. (1964). A philosophically plausible formal interpretation of intuitionistic logic. *Indagationes Mathematicae, 26*, 596–601.
Kamide, N. (2002). Relevance principle for substructural logics with mingle and strong negation. *Journal of Logic and Computation, 12*(6), 913–928.
Kashima, R. (2000). Completeness of implicational relevant logics. *Logic Journal of the IGPL, 8*(6), 761–785.
Kripke, S. A. (1959a). A completeness theorem in modal logic. *Journal of Symbolic Logic, 24*, 1–14.
Kripke, S. A. (1959b). The problem of entailment (abstract). *Journal of Symbolic Logic, 24*, 324.
Kripke, S. A. (1963). Semantical analysis of modal logic I. Normal modal propositional calculi. *Zeitschrift für mathematische Logik und Grundlagen der Mathematik, 9*, 67–96.
Kripke, S. A. (1965a). Semantical analysis of intuitionistic logic I. In J. N. Crossley & M. A. E. Dummett (Eds.), *Formal systems and recursive functions. Proceedings of the Eighth Logic Colloquium* (pp. 92–130). Amsterdam: North-Holland.
Kripke, S. A. (1965b). Semantical analysis of modal logic II. Non-normal propositional calculi. In J. W. Addison, L. Henkin, & A. Tarski (Eds.), *The theory of models* (pp. 206–220). Amsterdam: North-Holland.
Maksimova, L. L. (1973). A semantics for the system E of entailment. *Bulletin of the Section of Logic of the Polish Academy of Sciences, 2*, 18–21.
McCall, S. (1963). *Non-classical propositional calculi* (Unpublished doctoral dissertation). University of Oxford.
Meyer, R. K. (1966). *Topics in modal and many-valued logic* (Unpublished doctoral dissertation). Ann Arbor (UMI): University of Pittsburgh.
Meyer, R. K. (1968). *A characteristic matrix for RM*. (Manuscript. Published in Anderson and Belnap (1975), as §29.3, pp. 393–420.)

Meyer, R. K. (1970). E and S4. *Notre Dame Journal of Formal Logic, 11*(2), 181–199.
Meyer, R. K. (1976). Two questions from Anderson and Belnap. *Reports on Mathematical Logic, 7*, 71–86.
Meyer, R. K. (1985). *Proving semantical completeness "relevantly" for R* (Research Paper No. 23). Canberra: Logic Group, Department of Philosophy, Research School of Social Sciences, Australian National University.
Robles, G., Méndez, J. M., & Salto, F. (2010). A modal restriction of R-Mingle with the variable-sharing property. *Logic and Logical Philosophy, 19*, 341–351.
Segerberg, K. (1968). Propositional logics related to Heyting's and Johansson's. *Theoria, 34*, 26–61.
Slaney, J., Surendonk, T., & Girle, R. (1989). *Time, truth and logic* (Tech. Rep. Nos. TR-ARP-11/89). Canberra: Australian National University.
Urquhart, A. (1972). Semantics for relevant logic. *Journal of Symbolic Logic, 37*, 159–169.

# Chapter 7
# Modal, Fuzzy, …, Vanilla Fixpoint Theories of Truth: A Uniform Approach

**Melvin Fitting**

**Abstract** Kripke's work on modal logic has been immensely influential. It hardly needs remarking that this is not his only work. Here we address his pioneering applications of fixpoint constructions to the theory of truth, and related work by others. In his fundamental paper on this he explicitly described a modal version, applying a fixpoint construction world by world within a modal frame. This can certainly be carried out, and doubtless has been somewhere. Others have suggested a variety of other extensions such as using the unit interval as the underlying space of truth values, or using a four valued logic instead of three, or various combinations of these. When many similar formal constructions have been proposed, one naturally asks what is the common core. Is there some setting in which things can be proved once and for all, with the various specific proposals seen as applications of this common core. In fact this is the case, with bilattices providing the desired structure. Much of such a development has already appeared in some form or other. It is the purpose of this article to bring everything together, and also add a few things. We present general results that, more or less, have everything currently in the fixpoint literature as special cases. Fortunately this does not make things more complicated, since the underlying proofs are essentially the same as they have always been. It is just that they are being carried out in generality rather than in specificity.

**Keywords** Bilattice · First degree entailment · Fixed point theorems · Modal logic · Truth

M. Fitting (✉)
Graduate Center, City University of New York, Department of Philosophy (emeritus), 365 Fifth Avenue, New York City, USA
e-mail: melvin.fitting@gmail.com

© The Author(s), under exclusive license to Springer Nature Switzerland AG 2024

Y. Weiss and R. Birman (eds.), *Saul Kripke on Modal Logic*, Outstanding Contributions to Logic 30, https://doi.org/10.1007/978-3-031-57635-5_7

## 7.1 Introduction

In some areas of philosophy it is not uncommon to call on mathematical tools to help elucidate things. Saul Kripke is a master at this. Possible world semantics and fixed point theories of truth are two famous examples. They do not really solve philosophical problems themselves but, once having been introduced, they set the language and the conceptual structures we all use when these problems are debated. When philosophical disagreements about modality or about a truth predicate arise these mathematical constructions generally serve the role of ground bass above which discussions play themselves out. Personally, I have always had an interest in tools as such, and that is my subject now. This is most decidedly a technical paper.

The famous paper on truth, Kripke (1975), introduced a kind of *truth revision operator* with the idea that fixed points of the operator were natural candidates for ways an "is true" predicate might behave. Roughly, fixed points are interpretations of a truth predicate that can not be bettered using the information they contain. A similar idea appeared, essentially simultaneously, in Martin and Woodruff (1975). This paper built on Kleene's weak three valued logic, and used what are called *maximal* fixed points. Kripke took a broader view and examined three different logics: Kleene's strong three valued logic, Kleene's weak three valued logic, and supervaluations. In addition he investigated a range of different kinds of fixed points, with particular emphasis on the *least* one, but with the entire family playing a significant role. Both 1975 papers thought of classical logic as being at the heart of what they were doing, but Kripke noted that further extensions could be made.

> The present approach can be applied to languages containing modal operators. In this case we do not merely consider truth, but we are given, in the usual style of modal model theory, a system of possible worlds, and evaluate truth and $T(x)$ in each possible world. The inductive definition of the languages $\mathcal{L}_\alpha$ approximating to the minimal fixed point must be modified accordingly. We cannot give details here. (Kripke, 1975, 712-713)

The choice of a three valued logic, while natural, has some technical drawbacks. The underlying truth value structure is not that of a complete lattice, but rather of a complete partial ordering. This makes the mathematics somewhat more complicated—detailed machinery needed for the Kripke version can be found in Fitting (1986). Visser (1984) built on the idea that the Belnap-Dunn four-valued semantics for first degree entailment was an alternate, and better, setting for a fixed point approach. Here the structure is a complete lattice, and this allows the use of the remarkably simple Knaster-Tarski fixed point theorem. In addition the four-valued semantics includes within its structure both Kleene's strong three-valued logic and Graham Priest's logic of paradox, LP, which is also suitable for a Kripke-style development. There have also been fixed point versions in which the underlying truth value space was fuzzy. And this does not exhaust the list of similar developments.

In every variation or extension of the 1975 papers, results *similar* to those have been shown, by *similar* methods. Whenever a mathematician sees formal structures with similarities, about which similar things are established by similar methods, that mathematician will ask, "what is the common ground of this similarity?" That

common ground should be defined as an object of interest in its own right, and all these similar results should be proven just once, with all the rest as instances. For fixed point theories of truth this common ground is the *bilattice*. Why bilattices?

Bilattices are structures with two interconnected ordering relations. One of the relations can be used to give interpretations for logical connectives and quantifiers. These interpretations turn out to be monotonic with respect to the other order relation, and this fact can be used to show the existence of fixed points for truth revision operators. Kripke's work essentially took place in the simplest such example. He had only three truth values, roughly, true, false, and no-value-yet-assigned. This is actually not quite a full bilattice but it is the "consistent" part of the simplest one. In his setting one of the ordering relations was simply: false is less than true, with no-value-yet-assigned in between. Roughly speaking, this degree of truth ordering can be described by saying that an overall increase is either a decrease of falsehood or an increase of truth. The second ordering relation, the one for which monotonicity results obtain, is that no-value-yet-assigned is below both false and true, in effect a degree of information ordering.

Bilattices generalize the degree of truth and the degree of information orderings mentioned above. It will be seen that the bilattice family includes many natural examples that allow us to bring in modal logic, fuzzy logic, and other logics of interest as particular applications of general results. Some years ago I made use of bilattices in exactly this way in a series of papers: Fitting (1989, 1991, 1993, 1997, 2006), but things were rather spread out. The purpose of the present paper is to give a unified presentation, in full generality, with the items just mentioned as special cases, and with further applications provided as well.

The general structure of this paper is as follows.

In Sects. 7.2–7.4 the essential facts about bilattices are presented, important examples are given, and methods of constructing them are discussed. It is shown how propositional logic, first order logic, and modal logics can be semantically interpreted using bilattices as many valued truth value spaces.

In Sect. 7.5 several fixed point theorems are proved, all following from the Knaster-Tarski theorem. This is not quite the fixed point theorem used by Kripke, since his underlying algebraic structures were not complete lattices while ours are. Moving to complete lattices has the advantage of simplifying the mathematics, while still allowing the construction to proceed more or less as Kripke did it.

Section 7.6 adds a truth predicate to the language of logic we are using. The fundamental monotonicity theorems are then proved.

Kripke worked with three valued logics. These can be thought of as based on the consistent values of the four-valued Belnap-Dunn structure. In Sect. 7.7 the notion of consistent value is generalized to bilattices, along with the notions of exact (a generalization of the classical truth values), and anticonsistent (dual to that of consistent).

Section 7.8 finally brings all the preceding together, extending Kripke's treatments of fixed point truth predicates using generalizations of Kleene's strong and Kleene's weak three valued logics, and generalizations of supervaluations, as well as a brief discussion involving asymmetric logics.

Section 7.9 examines some more specialized fixed point work, generalizing the notions of intrinsic and maximal. It also brings in so-called alternating fixed points, which actually first arose in computer science, but which have significance here as well.

## 7.2 Bilattices

There is an important feature common to fixpoint theories of truth generally: two partial orderings are involved. There is a space of truth values, with its ordering on *degree of truth* (or falsehood). There is also some truth revision operator involved, for which one wants fixed points, so some notion of approximation to a fixed point is desired. This always involves an ordering on what might be called the *degree of information*. The structure common to fixpoint truth theories, then, has two partial orderings, truth and information, somehow interconnected—and this is a *bilattice*.

Bilattices can have various conditions imposed. What we need here are bilattices with quite a bit of structure. They should be *complete, infinitarily interlaced bilattices, with a negation and a conflation that commute*. This is a lot of terminology, but the underlying ideas are all quite natural and are explained below. But as background we need lattices, bounded, and complete, so let us start there.

**Definition 7.2.1** (Lattices) A *partially ordered* set is a a set with a binary relation, $\leq$, that is reflexive ($x \leq x$), transitive ($x \leq y$ and $y \leq z$ imply $x \leq z$), and anti-symmetric ($x \leq y$ and $y \leq x$ imply $x = y$).

A *lattice* is a partial ordering in which every pair of elements has a unique least upper bound (join) and a unique greatest lower bound (meet). (It follows that least upper and greatest lower bounds exist for every finite, non-empty set).

A *bounded* lattice is a lattice with a unique largest and a unique smallest element, usually called *top* and *bottom*.

A lattice is *complete* if every set (not just the finite ones) has a unique least upper bound and a unique greatest lower bound. (This implies boundedness, since the greatest lower bound of the empty set is top, and the least upper bound of the empty set is bottom.)

Now we move to our real subject, bilattices. A pre-bilattice is simply a structure with two orderings. Once conditions connecting the two orderings are postulated, it is standard to drop the 'pre' qualification.

**Definition 7.2.2** (Bilattices) Terminology is as follows.

Pre-Bilattice: A pre-bilattice is a structure $\mathcal{B} = \langle \mathcal{B}, \leq_t, \leq_k \rangle$ with two lattice orderings; informally $\leq_t$ is on degree of truth and $\leq_k$ is on degree of information.

Extreme Elements: It is common to assume each lattice ordering is bounded, with a top and a bottom. We do so here. The smallest and largest members with respect to the truth ordering $\leq_t$ will be denoted $\mathbf{f}$ and $\mathbf{t}$, and with respect to the information ordering $\leq_k$, $\bot$ and $\top$.

Bilattice Operations: For $x, y \in \mathcal{B}$ their meet with respect to $\leq_t$ is denoted $x \wedge y$ ("$\wedge$" is simply read "and"), the join is denoted $x \vee y$ ("$\vee$" is read "or"). With respect to $\leq_k$ the meet is $x \otimes y$ ("$\otimes$" is read "consensus"), and join is $x \oplus y$ ("$\oplus$" is read "gullability" or "accept all").

Interlaced: It follows from the definitions of least upper and greatest lower bounds, that each of the four bilattice operations is monotone with respect to its ordering. For example, if $x \leq_k x'$ and $y \leq_k y'$ then $x \otimes y \leq_k x' \otimes y'$. A bilattice is *interlaced* if each operation is monotone with respect to *both* orderings. As one example, $x \leq_t x'$ and $y \leq_t y'$ then $x \otimes y \leq_t x' \otimes y'$.

Negation: A bilattice has a *negation* if there is an involution $\neg : \mathcal{B} \to \mathcal{B}$ that reverses the truth ordering and leaves intact the information ordering. Thus $\neg\neg x = x$, and $x \leq_t y$ implies $\neg y \leq_t \neg x$, and $x \leq_k y$ implies $\neg x \leq_k \neg y$.

Conflation: A bilattice has a *conflation* if there is an involution $- : \mathcal{B} \to \mathcal{B}$ that reverses the information ordering and leaves intact the truth ordering. Thus $- - x = x$, and $x \leq_k y$ implies $-y \leq_k -x$, and $x \leq_t y$ implies $-x \leq_t -y$. Negation and conflation commute if always $\neg - x = -\neg x$.

Completeness: A bilattice is *complete* if both $\langle \mathcal{B}, \leq_t \rangle$ and $\langle \mathcal{B}, \leq_k \rangle$ are complete lattices. We write $\bigwedge S$ and $\bigvee S$ for the meet and join of set $S$ with respect to the $\leq_t$ ordering, and $\prod S$ and $\sum S$ for meet and join with respect to the $\leq_k$ ordering.

Infinitarily Interlaced: A bilattice is *infinitarily interlaced* if all meets and joins with respect to each ordering are monotone with respect to the other ordering, as well as to their own. As one example, suppose we have an indexing set $I$, both $\{a_i \mid i \in I\}$ and $\{b_i \mid i \in I\}$ are subsets of $\mathcal{B}$, and $a_i \leq_t b_i$ for each $i \in I$. Then $\prod_i a_i \leq_t \prod_i b_i$.

The conditions above are not all independent. Completeness implies the existence of tops and bottoms for both orderings, and infinitarily interlaced implies interlaced. For applications in other areas not all of the conditions may be required, but we will want all of them except sometimes conflation. It is a lot to keep saying that we have a pre-bilattice that is interlaced, has a negation and a conflation that commute, is complete, and is infinitarily interlaced. Consequently we specialize terminology here. When reading other papers, please check usage there, to avoid confusion.

**Terminology Convention 7.2.3** From now on a bilattice meeting all the conditions above, omitting conflation, will simply be referred to here as a *bilattice*. This is an abuse of terminology, to be sure, but it simplifies verbiage and should cause no general problems. By a *bilattice with conflation* we mean a bilattice in which there is a conflation operation that commutes with negation.

Thinking of a bilattice as a space of truth values, logical connectives for conjunction, disjunction, and negation can be interpreted using the truth operations of the bilattice. It should be no problem if we use "$\wedge$" for both the logical connective and for the bilattice operation, since context will sort things out. Similarly for "$\vee$" and "$\neg$". Quantifiers will be interpreted using the infinitary meet, "$\bigwedge$", and join, "$\bigvee$", operations of the truth ordering. The information operations will be used too, but for purposes other than directly interpreting logical operations.

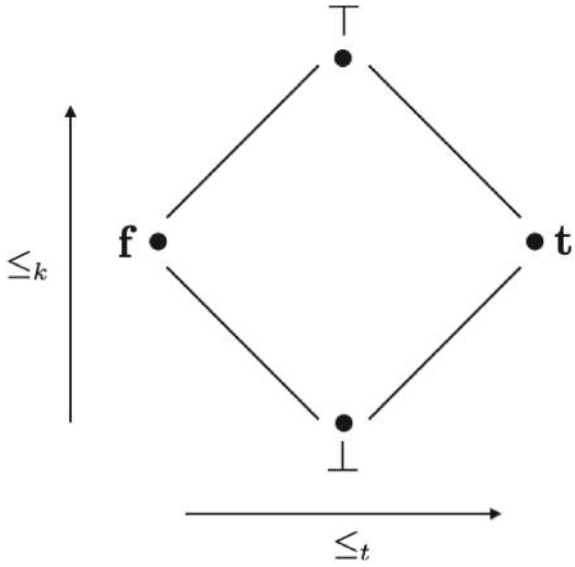



**Fig. 7.1** The Bilattice $\mathcal{FOUR}$

**Example 7.2.4** Figure 7.1 shows an example of a bilattice with conflation, the simplest one since it has only the four extreme values which every bilattice must have. It is certainly the most important example, and we refer to Arieli and Avron (1998) for some of the reasons why. It is standard to display bilattices using a double Hasse diagram. The truth ordering is from left to right, and the information ordering is from bottom upwards. The bilattice $\mathcal{FOUR}$ derives from Belnap and Dunn, see Belnap (1977), and its truth ordering provides the standard semantics for first degree entailment (which will not play a role here). It is actually an example of a *distributive* bilattice, that is, each operation distributes over all the others. For instance, $x \wedge (y \oplus z) = (x \wedge y) \oplus (x \wedge z)$. It is not hard to show that distributivity implies interlacing. The converse is not true. We will not mention distributivity further since interlacing has turned out to be the more important concept for bilattices. Since $\mathcal{FOUR}$ is interlaced it is trivially infinitarily interlaced, being finite. Negation is left-right symmetry. Conflation is vertical symmetry. We will revisit this example from time to time.

**Proposition 7.2.5** (De Morgan Laws) *In the following the left entries for are bilattices generally, and the right are for bilattices with conflation. For a set S of bilattice members we write* $\neg S$ *for* $\{\neg x \mid x \in S\}$*, and similarly for* $-S$*.*

$$\begin{array}{ll}
\neg(x \wedge y) = \neg x \vee \neg y & -(x \wedge y) = -x \wedge -y \\
\neg(x \vee y) = \neg x \wedge \neg y & -(x \vee y) = -x \vee -y \\
\neg(x \otimes y) = \neg x \otimes \neg y & -(x \otimes y) = -x \oplus -y \\
\neg(x \oplus y) = \neg x \oplus \neg y & -(x \oplus y) = -x \otimes -y \\
\neg \bigwedge S = \bigvee \neg S & -\bigwedge S = \bigwedge -S \\
\neg \bigvee S = \bigwedge \neg S & -\bigvee S = \bigvee -S \\
\neg \prod S = \prod \neg S & -\prod S = \sum -S \\
\neg \sum S = \sum \neg S & -\sum S = \prod -S
\end{array}$$

The following says the four extreme members of a bilattice are always tightly connected. It can be proved directly but here it is a special case of Proposition 7.5.8.

**Proposition 7.2.6** *In any bilattice:* $\mathbf{t} \otimes \mathbf{f} = \bot$, $\mathbf{t} \oplus \mathbf{f} = \top$, $\top \wedge \bot = \mathbf{f}$, *and* $\top \vee \bot = \mathbf{t}$.

## 7.3 Constructing Bilattices

There is a universal way of constructing bilattices, and also a function space construction that is not universal but is essential for what we will do. This does not exhaust the subject of bilattice constructions, but is sufficient for our needs here.

### *7.3.1 Bilattice Products*

We are about to present an intuitive and universal construction for bilattices. It entered the bilattice community, with various names, through Ginsberg (1988), Fitting (1990, 1991), Avron (1996), Pynko (2000). It turns out that, in fact, it was previously known using different terminology and in a different context, see Gargov (1999), Davey (2013). We are primarily interested in a special case here, because of our Terminology Convention 7.2.3, but we do mention more general versions. In the following, note the reversal involving the second component, for the $\leq_t$ ordering.

**Definition 7.3.1** (Bilattice Product) Let $L_1 = \langle L_1, \leq_1 \rangle$ and $L_2 = \langle L_2 \leq_2 \rangle$ be two lattices. The *bilattice product*, $L_1 \odot L_2$, is the structure $\langle L_1 \times L_2, \leq_t, \leq_k \rangle$ where

1. $\langle x_1, y_1 \rangle \leq_t \langle x_2, y_2 \rangle$ if $x_1 \leq_1 x_2$ and $y_2 \leq_2 y_1$,
2. $\langle x_1, y_1 \rangle \leq_k \langle x_2, y_2 \rangle$ if $x_1 \leq_1 x_2$ and $y_1 \leq_2 y_2$.

There is a natural intuition behind this. Think of $L_1$ as representing degree of evidence *for* some proposition, and $L_2$ as representing degree of evidence *against*. These need not be measured the same way. A value in a bilattice product separately records evidence for and evidence against. Degree of information (the $\leq_k$ ordering) goes up if we have the same or more evidence both for and against. Degree of truth (the $\leq_t$ ordering) goes up if evidence for goes up while evidence against goes down.

The bilattice product construction is quite versatile. In the following discussion we use $\sqcap_1$ and $\sqcup_1$ for meet and join of lattice $L_1$ and $\sqcap_2$ and $\sqcup_2$ for meet and join of lattice $L_2$. Likewise if the lattices are bounded, we use $0_1$ and $1_1$ for bottom and top of $L_1$, and similarly for $L_2$. If the lattices are complete we use $\bigsqcap_1$ and $\bigsqcup_1$ for the meet and join of arbitrary subsets of $L_1$ and similarly for $L_2$.

1. $L_1 \odot L_2$ is a lattice with respect to both orderings, with the following operations.

$$\begin{aligned}
\langle x_1, y_1\rangle \oplus \langle x_2, y_2\rangle &= \langle x_1 \sqcup_1 x_2, y_1 \sqcup_2 y_2\rangle \\
\langle x_1, y_1\rangle \otimes \langle x_2, y_2\rangle &= \langle x_1 \sqcap_1 x_2, y_1 \sqcap_2 y_2\rangle \\
\langle x_1, y_1\rangle \vee \langle x_2, y_2\rangle &= \langle x_1 \sqcup_1 x_2, y_1 \sqcap_2 y_2\rangle \\
\langle x_1, y_1\rangle \wedge \langle x_2, y_2\rangle &= \langle x_1 \sqcap_1 x_2, y_1 \sqcup_2 y_2\rangle
\end{aligned}$$

2. $L_1 \odot L_2$ always meets the interlacing conditions.
3. If $L_1$ and $L_2$ are bounded lattices $L_1 \odot L_2$ has extreme elements: $\langle 0_1, 0_2\rangle$ and $\langle 1_1, 1_2\rangle$ for the information ordering; $\langle 0_1, 1_2\rangle$ and $\langle 1_1, 0_2\rangle$ for the truth ordering.
4. If $L_1$ and $L_2$ are complete lattices then $L_1 \odot L_2$ is a complete bilattice, with the following infinitary operations.

$$\begin{aligned}
\sum S &= \langle \textstyle\bigsqcup_1 \{x \mid \langle x, y\rangle \in S \text{ where } y \in L_2\}, \bigsqcup_2 \{y \mid \langle x, y\rangle \in S \text{ where } x \in L_1\}\rangle \\
\prod S &= \langle \textstyle\bigsqcap_1 \{x \mid \langle x, y\rangle \in S \text{ where } y \in L_2\}, \bigsqcap_2 \{y \mid \langle x, y\rangle \in S \text{ where } x \in L_1\}\rangle \\
\bigvee S &= \langle \textstyle\bigsqcup_1 \{x \mid \langle x, y\rangle \in S \text{ where } y \in L_2\}, \bigsqcap_2 \{y \mid \langle x, y\rangle \in S \text{ where } x \in L_1\}\rangle \\
\bigwedge S &= \langle \textstyle\bigsqcap_1 \{x \mid \langle x, y\rangle \in S \text{ where } y \in L_2\}, \bigsqcup_2 \{y \mid \langle x, y\rangle \in S \text{ where } x \in L_1\}\rangle
\end{aligned}$$

5. If $L_1$ and $L_2$ are complete lattices then $L_1 \odot L_2$ is infinitarily interlaced.
6. If $L_1 = L_2 = L$, that is, if the two lattices are the same, then $L \odot L$ has a negation, $\neg\langle x, y\rangle = \langle y, x\rangle$.

Conflation needs some additional machinery.

**Definition 7.3.2** (De Morgan Lattice) A De Morgan lattice is a bounded distributive lattice (Definition 7.2.1) with a De Morgan involution operation $x \mapsto \overline{x}$ that is order reversing, $x \leq y$ implies $\overline{y} \leq \overline{x}$, and is an involution, $\overline{\overline{x}} = x$. Since we will never need distributivity, we call a structure a *non-distributive* De Morgan lattice if it meets all the conditions except possibly for the distributive laws.

It is easy to show De Morgan laws hold in any non-distributive De Morgan lattice. One more item is now added to the list above.

7. If $L_1 = L_2 = L$ is a non-distributive De Morgan lattice, then $L \odot L$ has a conflation, $-\langle x, y\rangle = \langle \overline{y}, \overline{x}\rangle$, and it commutes with negation, item 6.

We now have a simple way of constructing bilattices. It is completely general. For instance, consider items 2 and 3. In the converse direction it can be shown that every bilattice with extreme elements that meets the interlacing conditions is isomorphic to $L_1 \odot L_2$ where both $L_1$ and $L_2$ are bounded lattices. And so on. Thus we actually have full representation theorems.

### *7.3.2 Function Space Bilattices*

This is another way of constructing bilattices that will be of much use to us. It is not universal, unlike bilattice product, but it does help to unify several constructions that have appeared in the literature on fixed point theories of truth.

**Definition 7.3.3** (Function Space Bilattice) Let $\mathcal{B} = \langle \mathcal{B}, \leq_t, \leq_k\rangle$ be a bilattice, and $\mathcal{A}$ be a non-empty set. The function space $\mathcal{B}^{\mathcal{A}} = \langle \mathcal{B}^{\mathcal{A}}, \leq_t, \leq_k\rangle$ is a bilattice in which we take as domain the set consisting of all functions from $\mathcal{A}$ to $\mathcal{B}$, and with orderings that are pointwise, which means that for $f, g \in \mathcal{B}^{\mathcal{A}}$:

1. $f \leq_t g$ iff $f(x) \leq_t g(x)$ for all $x \in \mathcal{A}$,
2. $f \leq_k g$ iff $f(x) \leq_k g(x)$ for all $x \in \mathcal{A}$.

Notice that we have used $\leq_t$ and $\leq_k$ to denote partial orderings in both $\mathcal{B}$ and $\mathcal{B}^{\mathcal{A}}$. This keeps notation relatively simple, and context should always clarify which bilattice is meant. A function space bilattice inherits properties from the underlying bilattice. We leave verifications of the following to you.

1. $\mathcal{B}^{\mathcal{A}}$ is a lattice with respect to both orderings. For instance, the meet with respect to $\leq_k$ is easily verified to be the function given by $(f \oplus g)(x) = f(x) \oplus g(x)$, and similarly with the other meet and both joins.
2. $\mathcal{B}^{\mathcal{A}}$ is interlaced (since $\mathcal{B}$ is).
3. $\mathcal{B}^{\mathcal{A}}$ is bounded with respect to both orderings. For instance, the least member of $\mathcal{B}^{\mathcal{A}}$ in the pointwise $\leq_k$ ordering is the function that is identically $\bot$ on $\mathcal{A}$, where $\bot$ is the least member of $\mathcal{B}$ in the $\leq_k$ ordering on $\mathcal{B}$. Similarly for the other three extreme elements.
4. $\mathcal{B}^{\mathcal{A}}$ is a complete bilattice (since $\mathcal{B}$ is). If $S \subseteq \mathcal{B}^{\mathcal{A}}$, $\sum S$ is the function such that $(\sum S)(x) = \sum\{f(x) \mid f \in S\}$, and similarly for the other cases.
5. $\mathcal{B}^{\mathcal{A}}$ is infinitarily interlaced.
6. $\mathcal{B}^{\mathcal{A}}$ has a negation,and $(\neg f)(x) = \neg(f(x))$.
7. If $\mathcal{B}$ has a conflation that commutes with negation, the same is true of $\mathcal{B}^{\mathcal{A}}$,and $(-f)(x) = -(f(x))$.

If we have a space of truth values that constitutes a bilattice, such as $\mathcal{FOUR}$, the space of valuations in this bilattice, which is a function space, is actually a bilattice. It is here that most of the fixpoint machinery is actually used.

## 7.4 Bilattice Examples

We have already seen the most basic bilattice example, $\mathcal{FOUR}$, in Fig. 7.1, which is an easy bilattice product construction. Start with the simplest Boolean algebra with underlying space $\{0, 1\}$, and with $0 \leq 1$. Identifying 0 with falsehood and 1 with truth, the meet and join are those of classical propositional logic. There is a natural De Morgan involution, switch 0 and 1—this corresponds to classical negation. The bilattice product of this structure with itself is, isomorphically, $\mathcal{FOUR}$, with $\bot$ being $\langle 0, 0\rangle$, $\top$ being $\langle 1, 1\rangle$, **f** as $\langle 0, 1\rangle$ and **t** as $\langle 1, 0\rangle$. We now move on to several less familiar examples.

**Example 7.4.1** (*Bilattices of Sets of Agents*) Suppose we have two sets of agents, $L_1$ and $L_2$. Members of $L_1$ get to vote that a particular informal proof of some proposition is correct, or else offer no opinion. Members of $L_2$ play a similar role with respect to the correctness of some informally presented countermodel to the proposition. These are our 'for' and 'against' agents. The two kinds of agents could even be drawn from different populations. A natural lattice ordering relation is simply subset, so degree goes up if more agents join a group. We have no requirement that everybody speaks up on every proposition; indeed there might be no agents either for or against some proposition. We have no requirement that agents be consistent, some might belong to both groups and decide to accept both a proof and a countermodel. This can easily happen since both proofs and countermodels might be allowed to be informal, so there is the possibility of error or misunderstanding. At any rate, a member of the bilattice product $L_1 \odot L_2$ encodes what set of agents is for, and what set is against.

Degree of information increases if more agents join one or another (or both) of the sets. Degree of truth goes up if more join the set for or leave the set against. The bottom in the truth ordering is $\langle \emptyset, L_2\rangle$, nobody accepts an informal proof, everybody accepts a countermodel. Similarly for the other extreme bilattice values.

If the sets of agents, for and against, are drawn from the same population, that is if $L_1 = L_2$, then negation simply reverses the positions of those who are for and those who are against. Further, if $L_1 = L_2$ the appropriate De Morgan involution is relative compliment with respect to the entire set of agents. Then conflation in the bilattice product of agents replaces the set of agents for something with the set of those who were not against it, and the set against with the set of those who were not for.

$\mathcal{FOUR}$ can be thought of as an agent bilattice with a single agent. There are two agent subsets, none and all, and these can be identified with the members of $\{0, 1\}$.

**Example 7.4.2** (*Bilattice $\mathcal{NINE}$*) This example is similar to $\mathcal{FOUR}$, but now start with the space $\{0, \frac{1}{2}, 1\}$ with the standard numerical ordering. The bilattice product of this with itself is in Fig. 7.2. Properties are similar to those of $\mathcal{FOUR}$. It can be thought of as a bilttice built on Kleene's strong three valued logic, or Priest's logic of paradox.

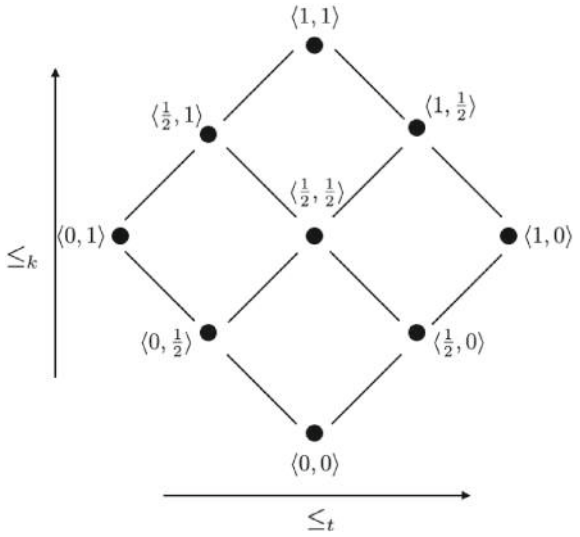



**Fig. 7.2** The Bilattice $\mathcal{NINE}$

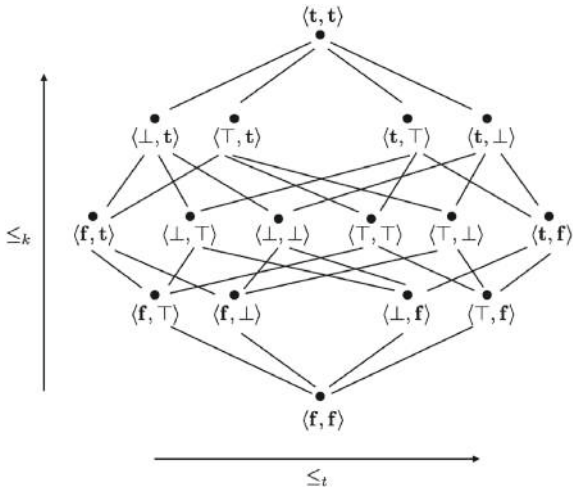
**Example 7.4.3** (*Bilattice* $\mathcal{SIXTEEN}$) Start with the lattice from $\mathcal{FOUR}$ just using the $\leq_t$ ordering. The bilattice product of this with itself is shown in Fig. 7.3. It has features of interest that will be discussed from time to time.

**Example 7.4.4** (*A Fuzzy Bilattice*) Examples so far have been finite. All have been complete bilattices, simply because they are finite. Now we look at a 'fuzzy' example. Consider the closed unit real interval, [0, 1], with the usual numerical ordering. Form the bilattice product of this with itself, so a bilattice truth value is a pair $\langle x, y\rangle$ where $x$ can be thought of as a degree of belief in some proposition, and $y$ as a corresponding degree of doubt. (Think of belief and doubt as being independently arrived at.) Loosely, one can think of it as $\mathcal{FOUR}$, but 'filled in'. This is a bilattice that is complete, not because it is finite, but because the unit interval ordering is a complete one. It is also a bilattice with conflation, where $-\langle x, y\rangle = \langle 1 - y, 1 - x\rangle$.

**Example 7.4.5** (*Propositional Valuations*) The previous examples made use of bilattice products. Function space bilattices also arise quite naturally. Let $\mathcal{A}$ be a set of propositional atoms, from which formulas of propositional language $\mathcal{L}$ are built using connectives $\wedge$, $\vee$, and $\neg$. Let our space of truth values be a bilattice $\mathcal{B}$, for instance any of the examples discussed above. A *propositional valuation* is any mapping from the set of atoms $\mathcal{A}$, to the bilattice $\mathcal{B}$, an arbitrary member of $\mathcal{B}^{\mathcal{A}}$. This gives us a bilattice, $\langle \mathcal{B}^{\mathcal{A}}, \leq_t, \leq_k\rangle$, using pointwise orderings as described in Sect. 7.3.2. For propositional valuations $v$ and $w$ we have $v \leq_k w$ if $v(P) \leq_k w(P)$ for every propositional letter $P$, and similarly for $\leq_t$.

One could also consider a function space bilattice $\mathcal{B}^{\mathcal{L}}$, where $\mathcal{L}$ is the entire set of propositional formulas, and not just the atomic ones. This is not of much interest as a

**Fig. 7.3** The Bilattice $\mathcal{SIXTEEN}$




whole, but it does have an important subspace. Each propositional valuation extends to all formulas in a unique way so that the following conditions are met, using the truth operations from the bilattice $\mathcal{B}$ in place of conventional truth tables.

$$\begin{aligned} v(X \wedge Y) &= v(X) \wedge v(Y) \\ v(X \vee Y) &= v(X) \vee v(Y) \\ v(\neg X) &= \neg v(X). \end{aligned}$$

On the left we have operation symbols of the language $\mathcal{L}$, and on the right operations from the truth ordering in the bilattice $\mathcal{B}$. Note that we have used the same notation for valuations in $\mathcal{B}^{\mathcal{A}}$ and for their extensions in $\mathcal{B}^{\mathcal{L}}$, relying on context to clarify which is meant. Valuations, extended as above to mappings on all formulas, constitute an obvious natural subspace of the bilattice $\mathcal{B}^{\mathcal{L}}$. Of the two bilattice orderings restricted to this subspace, $\leq_t$ is not interesting (because of the behavior of negation), but $\leq_k$ most decidedly is of interest.

**Proposition 7.4.6** (Propositional Monotonicity) *For propositional valuations* $v, w \in \mathcal{B}^{\mathcal{A}}$,

$$v \leq_k w \text{ in } \mathcal{B}^{\mathcal{A}} \text{ if and only if } v \leq_k w \text{ in } \mathcal{B}^{\mathcal{L}}.$$

Stated more explicitly, for valuations $v$ and $w$, extended to all formulas as above, $v(P) \leq_k w(P)$ for all propositional atoms $P$ if and only if $v(X) \leq_k w(X)$ for all formulas $X$. The implication from right to left is trivial. The implication from left to right is shown by a simple induction on formula complexity, and we leave it to you.

**Example 7.4.7** (*Quantificational Valuations*) Let formal language $\mathcal{L}$ now be first-order, with relation symbols but no constant or function symbols (for convenience

only). Assume language $\mathcal{L}$ has both universal and existential quantifiers. Let $\mathcal{D}$ be some non-empty domain, which we think of quantifiers as ranging over. Doing things a little differently from the standard (but following Smullyan, 1968) rather than allowing free variables to occur in formulas and working with a function assigning members of the domain to those free variables, we simply allow members of the domain to occur directly in formulas. By a $\mathcal{D}$ sentence we mean a formula of $\mathcal{L}$ in which all free variable occurrences have been replaced with members of $\mathcal{D}$—we write $\mathcal{L}(\mathcal{D})$ for the set of all $\mathcal{D}$ sentences. Atomic $\mathcal{D}$ sentences are of the form $P(a_1, \ldots, a_n)$ where $P$ is a relation symbol and $a_1, \ldots, a_n$ are members of $\mathcal{D}$. We call the set of atomic $\mathcal{D}$ sentences $\mathcal{A}(\mathcal{D})$.

Let $\mathcal{B}$ be a bilattice of intended truth values. A *quantificational valuation* in $\mathcal{B}$ is a mapping from atomic $\mathcal{D}$ sentences to members of $\mathcal{B}$, a member of function space bilattice $\mathcal{B}^{\mathcal{A}(\mathcal{D})}$. Quantificational valuation $v$ can be extended to all $\mathcal{D}$ sentences, that is, to a mapping in $\mathcal{B}^{\mathcal{L}(\mathcal{D})}$. Propositional connectives are handled exactly as in Example 7.4.5. The quantificational cases are as follows.

$$v(\forall x F(x)) = \bigwedge \{v(F(d)) \mid d \in \mathcal{D}\}$$
$$v(\exists x F(x)) = \bigvee \{v(F(d)) \mid d \in \mathcal{D}\}$$

Proposition 7.4.6 extends to the following, whose proof we again leave to you.

**Proposition 7.4.8** (Quantificational Monotonicity) *For quantificational valuations* $v, w \in \mathcal{B}^{\mathcal{A}(\mathcal{D})}$,

$$v \leq_k w \text{ in } \mathcal{B}^{\mathcal{A}(\mathcal{D})} \text{ if and only if } v \leq_k w \text{ in } \mathcal{B}^{\mathcal{L}(\mathcal{D})}.$$

**Example 7.4.9** (*Modal Valuations*) Both propositional and quantified languages extend to include modal operators. We only discuss the quantified version, and we assume we have constant domain models. Modifying the work to allow varying domain is not hard, but the essential ideas are more easily discerned in the present, simpler, setting. We start with the first-order machinery of Example 7.4.7, a bilattice $\mathcal{B}$, a domain $\mathcal{D}$ of quantification, and a set $\mathcal{A}(\mathcal{D})$ of atomic $\mathcal{D}$ sentences incorporating members of $\mathcal{D}$ directly. The first-order language is extended with modal operators $\Box$ and $\Diamond$ in the usual way. We continue to use $\mathcal{L}(\mathcal{D})$ for the set of $\mathcal{D}$ sentences, but now allowing modal operators. As new semantic machinery suppose we have a modal *frame*, $\mathcal{F} = \langle \mathcal{G}, \mathcal{R} \rangle$, where $\mathcal{G}$ is a non-empty set of possible worlds and $\mathcal{R}$ is a binary accessibility relation. We assume the basics of possible world semantics for modal logics are understood.

Form the function space bilattice $\mathcal{B}^{\mathcal{G}}$, with pointwise orderings. Since $\mathcal{B}^{\mathcal{G}}$ is a bilattice, $(\mathcal{B}^{\mathcal{G}})^{\mathcal{A}(\mathcal{D})}$ is also a bilattice, of the kind discussed in Example 7.4.7 except that the underlying bilattice is now itself a function space bilattice. A member of this bilattice is a valuation function $v : \mathcal{A}(\mathcal{D}) \to (\mathcal{G} \to \mathcal{B})$ that assigns to each atomic $\mathcal{D}$ sentence, and each possible world, a truth value in the bilattice $\mathcal{B}$. We call members of this *quantificational modal valuations*.

Suppose $v \in (\mathcal{B}^{\mathcal{G}})^{\mathcal{A}(\mathcal{D})}$; in Example 7.4.5 we said how to extend $v$ from the atomic level, adding cases involving propositional connectives, and this carries over to here.

We sketch how it gives us the appropriate behavior for propositional connectives in the modal setting as well. Suppose that $v \in (\mathcal{B}^{\mathcal{G}})^{\mathcal{L}(\mathcal{D})}$ meets the conditions for $\wedge$, $\vee$, and $\neg$ from Example 7.4.5. Then $v(X \wedge Y) = v(X) \wedge v(Y)$. But $v$ maps to $\mathcal{B}^{\mathcal{G}}$, so $v(X \wedge Y)$ is a function from $\mathcal{G}$ to $\mathcal{B}$, and similarly for $v(X)$ and $v(Y)$. Then for a possible world $\Gamma \in \mathcal{G}$, $v(X \wedge Y)(\Gamma) = (v(X) \wedge v(Y))(\Gamma)$. The $\wedge$ on the right is an operator of $\mathcal{B}^{\mathcal{G}}$, and behaves pointwise, so $(v(X) \wedge v(Y))(\Gamma) = v(X)(\Gamma) \wedge v(Y)(\Gamma)$. Thus we have $v(X \wedge Y)(\Gamma) = v(X)(\Gamma) \wedge v(Y)(\Gamma)$. This is the usual way conjunction is understood in a possible world model. Disjunction and negation are similar.

The quantificational cases are as in Example 7.4.7, and here too we inherit appropriate world by world behavior automatically. For instance, we have the following: $v(\forall x \varphi(x))(\Gamma) = \bigwedge\{v(\varphi(d))(\Gamma) \mid d \in \mathcal{D}\}$, and similarly for $\exists$.

What must be added is extension conditions to cover the modal operators. Recall, we are working with respect to a frame $\mathcal{F} = \langle \mathcal{G}, \mathcal{R} \rangle$. We now add the following, for each $\Gamma \in \mathcal{G}$, each $v \in (\mathcal{B}^{\mathcal{G}})^{\mathcal{A}(\mathcal{D})}$ and each $\mathcal{D}$ sentence $X$.

$$v(\Box X)(\Gamma) = \bigwedge\{v(X)(\Delta) \mid \Delta \in \mathcal{G} \text{ and } \Gamma \mathcal{R} \Delta\}$$
$$v(\Diamond X)(\Gamma) = \bigvee\{v(X)(\Delta) \mid \Delta \in \mathcal{G} \text{ and } \Gamma \mathcal{R} \Delta\}$$

We have covered all cases for extending a quantificational modal valuation $v \in (\mathcal{B}^{\mathcal{G}})^{\mathcal{A}(\mathcal{D})}$ to $v \in (\mathcal{B}^{\mathcal{G}})^{\mathcal{L}(\mathcal{D})}$, allowing modal operators. It would be a good exercise to show that $v(\neg\Box\neg X)(\Gamma) = v(\Diamond X)(\Gamma)$ for every $\mathcal{D}$ sentence $X$ and every $\Gamma \in \mathcal{G}$.

Propositions 7.4.6 and 7.4.8 now easily extend to the following.

**Proposition 7.4.10** (Modal Monotonicity) *Using a modal first-order language $\mathcal{L}$ and a frame $\mathcal{F} = \langle \mathcal{G}, \mathcal{R} \rangle$, for modal valuations $v, w \in (\mathcal{B}^{\mathcal{G}})^{\mathcal{A}(\mathcal{D})}$,*

$$v \leq_k w \text{ in } (\mathcal{B}^{\mathcal{G}})^{\mathcal{A}(\mathcal{D})} \text{ if and only if } v \leq_k w \text{ in } (\mathcal{B}^{\mathcal{G}})^{\mathcal{L}(\mathcal{D})}.$$

It is clear that a rich variety of bilattices is available. $\mathcal{FOUR}$ and its various subspaces is often the setting for fixed point theories of truth but as we will see, essentially all the work applies quite generally. We suggest that Example 7.4.1 involving sets of agents is worth considering as a possible setting for thinking about the mutual understanding of a truth predicate in a community. This might be particularly interesting when the language involves modalities, perhaps even multimodalities, with a knowledge modal operator for each agent. Combining modal models with fuzzy bilattices, Example 7.4.4, is worth investigating. We leave such things to others.

## 7.5 Some Fixed Point Theorems

The application of fixed point theorems to theories of truth originated in Kripke (1975) and Martin and Woodruff (1975), but details are quite different between the two papers. Kripke used a *least* fixed point construction while Martin and Woodruff relied on *maximal* fixed points. Neither worked in a full bilattice setting, but rather in

a substructure of $\mathcal{FOUR}$ that we will be generalizing here. In a bilattice framework the simple and elegant fixed point result of Knaster and Tarski is the one that is most pertinent, and is what we will use here. This has, for its setting, a complete lattice, see Definition 7.2.1. In this section we begin with lattices, then move to bilattices. On the one hand we do not need the full Knaster-Tarski theorem, which has to do with the structure of the entire set of fixed points of a monotone mapping. All we need are greatest and least fixed points, and we skip the rest. On the other hand we will need both the original version and some simple generalizations.

**Definition 7.5.1** Let $\langle L, \leq \rangle$ be a lattice. A mapping $f : L \to L$ is *monotone* if it is order preserving, that is, if $x \leq y$ then $f(x) \leq f(y)$. A member $x \in L$ is: a *fixed point* of $f$ if $f(x) = x$, a *pre-fixed point* if $f(x) \leq x$, and a *post-fixed point* if $x \leq f(x)$.

**Proposition 7.5.2** (Knaster-Tarski) *Let $f$ be a monotone mapping on a complete lattice. Then:*

1. *$f$ has a smallest and a greatest fixed point.*
2. *If $a$ is a post-fixed point of $f$ then there is a fixed point of $f$, least above $a$.*
3. *If $b$ is a pre-fixed point of $f$ then there is a fixed point of $f$, greatest below $b$.*

The Knaster-Tarski theorem has two quite different proofs. One is short, self-contained, and somewhat magical; the other is substantially longer but provides more intuition. It is not self-contained, making use of facts from set theory concerning ordinal numbers. We sketch the longer proof, summarizing things properly belonging to set theory. We begin with ordinal numbers.

***Proof*** (*Sketch*) The sequence of ordinals starts with the natural numbers, 0, 1, 2, …, followed by the smallest infinite ordinal, $\omega$, which is followed by $\omega + 1$, $\omega + 2$ and so on. After these come the infinite ordinal, $\omega \cdot 2$, then $\omega \cdot 2 + 1$, $\omega \cdot 2 + 2$, and so on (a phrase that covers a lot). The ordinal $\omega$ is the first *limit* ordinal, $\omega \cdot 2$ is the second, with infinitely many limit ordinals beyond. Ordinals divide into three groups: successor ordinals, which can be written as $\alpha + 1$ where $\alpha$ is the immediate predecessor; limit ordinals, with predecessors but no immediate one; and 0, the only ordinal that is neither a limit ordinal nor a successor ordinal. The sequence of ordinals is not itself a set and so cannot be placed in a 1–1 correspondence with any set.

We begin with part 2. Let $\langle L, \leq \rangle$ be a complete lattice (assumed to be a set), with $f : L \to L$ a monotone function, and $a$ a post-fixed point of $f$, that is $a \leq f(a)$. We sketch why there is a fixed point for $f$ above $a$, least among fixed points above $a$.

Use transfinite recursion to define a mapping $\varphi$ from the class of ordinals to $L$, call it an *ordinal sequence*.

$$\begin{aligned} \varphi(0) &= a \text{ a post-fixed point of } f \\ \varphi(\alpha + 1) &= f(\varphi(\alpha)) \text{ for a successor ordinal } \alpha + 1 \\ \varphi(\lambda) &= \bigvee \{\varphi(\alpha) \mid \alpha < \lambda\} \text{ for a limit ordinal } \lambda \end{aligned}$$

The key fact about $\varphi$ is that it is monotonic, in the sense that if $\alpha$ and $\beta$ are ordinals and $\alpha \leq \beta$, then in lattice $L$ we have $\varphi(\alpha) \leq \varphi(\beta)$. The proof of this is set theoretic and a bit technical so we omit it. It follows that the sequence $\varphi(\alpha)$ increases in $L$ as ordinal $\alpha$ increases. That increase cannot be strict because otherwise we would be pairing the ordinals up in a 1–1 fashion with a subset of $L$, but the collection of ordinals cannot be paired up with the members of a set. Then at some point we must have $\varphi(\alpha) = \varphi(\alpha + 1)$. Let $\alpha_0$ be the least ordinal for which this happens. Then $\varphi(\alpha_0)$ is a fixed point of $f$ because $f(\varphi(\alpha_0)) = \varphi(\alpha_0 + 1) = \varphi(\alpha_0)$. Further, $\varphi(\alpha_0)$ is the *least* fixed point for $f$ above $a$, because if $F$ is any fixed point of $f$ above $a$ then $\varphi(\alpha) \leq F$ for all ordinals, so in particular, $\varphi(\alpha_0) \leq F$. This is shown by transfinite induction on $\alpha$, and again we omit the set theoretic proof.

Part 3 has a similar proof, but with everything dualized.

Part 1 now follows since $\bot$ is a post-fixed point of $f$ with every member of the lattice above it, and dually $\top$ is a pre-fixed point with every member below it. □

We not only have the existence of least and greatest fixed points in a complete lattice, but a way of proving that they have certain properties. Standard terminology: in a partial ordering two members $x$ and $y$ are called *comparable* if $x \leq y$ or $y \leq x$, and a subset is a *chain* if any two members are comparable.

**Definition 7.5.3** Let $\langle L, \leq \rangle$ be a complete lattice, $f$ be a monotone function on $L$, $a$ be a post-fixed point of $f$, and $P$ be a subset of $L$. We say $P$ is *$f$-inductive starting at $a$* if:

1. $a \in P$;
2. If $x \in P$ then $f(x) \in P$;
3. $P$ is *closed under chain sups*, meaning that if $S \subseteq P$ and $S$ is a chain then $\bigvee S \in P$.

**Corollary 7.5.4** *If $f$ is a monotone function on complete lattice $L$, $a$ is a post-fixed point of $f$, and $P \subseteq L$ is $f$-inductive starting at $a$, then the least fixed point of $f$ above $a$ is in $P$.*

***Proof*** Here is the idea. Using the hypotheses it is not hard to show all members of the ordinal sequence constructed in the proof of Proposition 7.5.2 are in $P$, so then the least fixed point of $f$ above $a$ is in $P$, since it is in the ordinal sequence. □

We note without proof that the Corollary dualizes to greatest fixed points.

There is a variation on the Knaster-Tarski result that we will need. Instead of finding a fixed point for a function, one seeks *two* values between which it oscillates. This requires not monotonicity, but *anti-monotonicity*.

**Definition 7.5.5** Let $\langle L, \leq \rangle$ be a lattice. A mapping $f : L \to L$ is *anti-monotone* if it is order reversing, $x \leq y$ implies $f(y) \leq f(x)$. An *alternating fixpoint pair* for $f$ is a pair $x$ and $y$ such that $f(x) = y$ and $f(y) = x$. An alternating fixpoint pair is *extremal* if any other alternating fixpoint pair is between them.

**Proposition 7.5.6** *Every anti-monotone mapping on a complete lattice has an extremal alternating fixpoint pair.*

***Proof*** This can be proved directly, but there is a simple derivation from Proposition 7.5.2. Let $\langle L, \leq\rangle$ be a complete lattice and $f : L \to L$ be anti-monotone. The function $f^2(x) = f(f(x))$ is obviously monotone, and so has fixed points; call the set of them $\mathcal{F}$. Since $f^2$ has a least and a greatest fixed point then $\mathcal{F}$ has a least member, $s$, and a greatest, $S$. We show these are an extremal alternating fixpoint pair for $f$.

We first show that $f$ maps $\mathcal{F}$ to $\mathcal{F}$. For, suppose $x \in \mathcal{F}$; then $f^2(f(x)) = f(f^2(x)) = f(x)$, so $f(x)$ is a fixed point of $f^2$, and hence in $\mathcal{F}$.

Since $s \in \mathcal{F}$ then $f(s) \in \mathcal{F}$. We show $f(s)$ is largest in $\mathcal{F}$, and hence is $S$. Let $x$ be an arbitrary member of $\mathcal{F}$. Then $f(x) \in \mathcal{F}$, so $s \leq f(x)$. By anti-monotonicity, $f(f(x)) \leq f(s)$, so $x \leq f(s)$. Since $x$ was arbitrary, $f(s) = S$. Then also $f(S) = f(f(s)) = s$, so $f$ alternates between $s$ and $S$.

Finally $s$ and $S$ are extremal. Suppose $x$ and $y$ are an alternating fixpoint pair for $f$. Then $f^2(x) = f(f(x)) = f(y) = x$ so $x$ is a fixpoint of $f^2$. Similarly for $y$. Then $x$ and $y$ are in $\mathcal{F}$ and so are between $s$ and $S$. □

We now move on to our main subject, bilattices. Recall we understand the term to include completeness with respect to each lattice order, as well as infinitary interlacing conditions connecting the orderings. We begin with a technical lemma whose proof is a prime example of interlacing conditions at work. It is essential for the subsequent results.

**Lemma 7.5.7** (From Avron, 1996) *In a bilattice $\mathcal{B}$:*

1. *For any $a \leq_t b$ we have $a \leq_t x \leq_t b$ if and only if $a \otimes b \leq_k x \leq_k a \oplus b$;*
2. *For any $a \leq_k b$ we have $a \leq_k x \leq_k b$ if and only if $a \wedge b \leq_t x \leq_t a \vee b$.*

***Proof*** We show the first item; the second is similar. Assume throughout that $a \leq_t b$.

1. left-right: Suppose $a \leq_t x \leq_t b$. By interlacing, $a \otimes (a \otimes b) \leq_t x \otimes (a \otimes b) \leq_t b \otimes (a \otimes b)$, or $a \otimes b \leq_t x \otimes (a \otimes b) \leq_t a \otimes b$. Then $x \otimes (a \otimes b) = a \otimes b$, and so $a \otimes b \leq_k x$. Similarly $x \leq_k a \oplus b$ by an argument using $\oplus$ instead of $\otimes$.
2. right-left: Suppose $a \otimes b \leq_k x \leq_k a \oplus b$. By interlacing, $a \wedge (a \otimes b) \leq_k a \wedge x \leq_k a \wedge (a \oplus b)$. Since $a \leq_t b$, by interlacing $a \otimes a \leq_t a \otimes b$, or $a \leq_t a \otimes b$, and hence $a \wedge (a \otimes b) = a$. Similarly $a \wedge (a \oplus b) = a$. Combining all this, $a = a \wedge (a \otimes b) \leq_k a \wedge x \leq_k a \wedge (a \oplus b) = a$. Then $a \wedge x = a$, and hence $a \leq_t x$. There is a similar proof that $x \leq_t b$. □

**Proposition 7.5.8** *Suppose $\langle \mathcal{B}, \leq_t, \leq_k\rangle$ is a bilattice, and $f : \mathcal{B} \to \mathcal{B}$ is a mapping that is monotone in* both *orderings. By the Knaster-Tarski theorem $f$ has a least and greatest fixed point with respect to each ordering. We use $p_t$, $P_t$ for smallest and biggest fixed point with respect to $\leq_t$, and $p_k$, $P_k$ for smallest and biggest with respect to $\leq_k$. The following hold: $P_t \otimes p_t = p_k$, $P_t \oplus p_t = P_k$, $P_k \wedge p_k = p_t$, $P_k \vee p_k = P_t$.*

***Proof*** $p_t$ and $P_t$ are least and greatest fixed points with respect to $\leq_t$, and $p_k$ is a fixed point, so $p_t \leq_t p_k \leq_t P_t$. By Lemma 7.5.7, $p_t \otimes P_t \leq_k p_k \leq_k p_t \oplus P_t$. Since $p_k$ is least fixed point under $\leq_k$, and $p_t$ and $P_t$ are fixed points, $p_k \leq_k p_t$ and $p_k \leq_k P_t$, so $p_k \leq_k p_t \otimes P_t$. Combining, $P_t \otimes p_t = p_k$. The other three are similar. □

As a special case of Proposition 7.5.8, let $f$ be the identity map on $\mathcal{B}$. Obviously this is monotone in both bilattice orderings, and everything is a fixed point, so the least and greatest fixed points under $\leq_t$ are **f** and **t**, while least and greatest under $\leq_k$ are $\bot$ and $\top$. Now Proposition 7.2.6 is an immediate consequence.

A version of Proposition 7.5.8 that we will need in Sect. 7.9.3 can now be obtained for the alternating case.

**Proposition 7.5.9** *Suppose* $\langle \mathcal{B}, \leq_t, \leq_k \rangle$ *is a bilattice and* $f : \mathcal{B} \to \mathcal{B}$ *is a mapping that is monotone with respect to* $\leq_k$ *but anti-monotone with respect to* $\leq_t$. *Let* $p_k$ *and* $P_k$ *be the least and greatest fixed points with respect to* $\leq_k$, *and let* $p_t$ *and* $P_t$ *be the smaller and the larger of the extremal alternating fixpoint pair with respect to* $\leq_t$. *The following hold:* $P_t \otimes p_t = p_k$, $P_t \oplus p_t = P_k$. $P_k \wedge p_k = p_t$, $P_k \vee p_k = P_t$.

***Proof*** Very simply, the mapping $f^2$ is monotone with respect to both orderings, and now use Proposition 7.5.8. □

## 7.6 Adding a Truth Predicate

We want to have a formal language that can speak about its own syntax. There are several ways of doing this; we use the traditional one, with an arithmetic language that talks about itself via Gödel numbering. Details are not particularly important. We take advantage of a point that Kripke has emphasized, and use relation symbols for addition and multiplication instead of function symbols. This has the nice consequence that every number has a unique term in the language that names it. In the following discussion we use symbols informally as names for themselves.

From now on assume $\mathcal{L}$ is a first-order language with a constant symbol 0 and a unary function symbol for successor—we write successor of $t$ as $t^+$. The only terms of the language are the open ones, $x^{+\cdots+}$ where $x$ is a variable, and the closed ones, $0^{+\cdots+}$, *numerals*. The number of successor symbols is allowed to be 0. We also assume we have a binary relation symbol $=$, and two ternary relation symbols $A$ and $M$, intended to represent addition and multiplication.

First-order formulas are built up using propositional connectives, $\wedge$, $\vee$, $\neg$, and quantifiers, $\forall$, $\exists$, in the standard way and, if modality is being considered, also using $\Box$ and $\Diamond$. When evaluating formulas we will make use of the standard model of arithmetic to assign meaning to the non-logical symbols, and to supply the range for the quantifiers. It is possible to have additional symbols to represent 'real world' things or relations, but this adds complexity without providing any fundamental new insights, so we avoid it here. Truth values for sentences of $\mathcal{L}$ will be in a bilattice, whose choice can be quite arbitrary. Keep $\mathcal{FOUR}$ in mind as an example.

In addition to the machinery described above, $\mathcal{L}$ has a unary predicate symbol, $\mathsf{T}$, intended to represent *is true*. Using it $\mathcal{L}$ can "talk about" the truth of its own sentences, using Gödel numbering. We ignore the details of Gödel numbering except that, for convenience, we assume every sentence (closed formula) has a Gödel number, and every number is the Gödel number of some sentence of $\mathcal{L}$. Numerals, 0, $0^+$, $0^{++}$, …, provide a unique representation for all numbers and so, via Gödel numbering, of every sentence. Let $X$ be a sentence of $\mathcal{L}$; by $\vDash X^{\urcorner}$ we mean the numeral representing the Gödel number of $X$. If $\mathsf{T}$ is to act like a truth predicate we should have equivalence of $X$ and $\mathsf{T}(\vDash X^{\urcorner})$ for every sentence $X$, that is, they should always have the same value in the bilattice that supplies our truth values. Since we have enough machinery to construct a liar sentence asserting its own non-truth, this can't be done if we use classical logic. Hence the historic move to three-valued logics, and here to bilattices.

Examples 7.4.7 and 7.4.9 looked at bilattices and first-order languages. In those examples we allowed members of domains to appear in sentences, as a convenient way of handling satisfiability. We no longer need to do this. We have numerals in $\mathcal{L}$, the natural numbers are the only domain we are interested in, and so we have names in $\mathcal{L}$ for the entire domain. Further, since arithmetic machinery will always be understood as if in the standard model for arithmetic, it is only the truth predicate $\mathsf{T}$ that can vary. So valuations will not assign values to all atomic sentences, but only to those involving $\mathsf{T}$.

**Definition 7.6.1** ($\mathsf{T}$ Valuation) $\mathcal{A}(\mathsf{T})$ is the set of atomic sentences of the form $\mathsf{T}(0^{++\cdots+})$ in the language $\mathcal{L}$. For any bilattice $\mathcal{B}$, a $\mathsf{T}$ *valuation in* $\mathcal{B}$ is a member of the function space bilattice $\mathcal{B}^{\mathcal{A}(\mathsf{T})}$ (Definition 7.3.3),

A $\mathsf{T}$ valuation in $\mathcal{B}$ simply assigns a truth value in $\mathcal{B}$ to every atomic sentence involving $\mathsf{T}$. Example 7.4.9 illustrates how the definition above covers the modal case, since the bilattice to which $\mathsf{T}$ valuations map can be of the form $\mathcal{B}^{\mathcal{G}}$ where $\mathcal{G}$ is the set of possible worlds of a modal frame. We will assume that if a language $\mathcal{L}$ includes modal operators, then the underlying bilattices are of this form.

Examples 7.4.7 and 7.4.9 showed how valuations extend from the atomic level to *all* sentences, that is, to members of the function space bilattice $\mathcal{B}^{\mathcal{L}}$. But in Kripke (1975) *three different* extensions were discussed, these have generalizations here, and others are possible. The following allows us flexibility while placing minimal requirements on extensions. Specific examples come in Sect. 7.8.

**Definition 7.6.2** (Valuation Extension) A *valuation extension* in bilattice $\mathcal{B}$ is a mapping $e : \mathcal{B}^{\mathcal{A}(\mathsf{T})} \to \mathcal{B}^{\mathcal{L}}$, where $\mathcal{L}$ is understood to be the set of *sentences* of our formal language. For $v \in \mathcal{B}^{\mathcal{A}(\mathsf{T})}$ we write the image under $e$ as $v^e$.

Our goal is to show that for several intuitively acceptable valuation extensions $e$ there are valuations $v \in \mathcal{B}^{\mathcal{A}(\mathsf{T})}$ such that for each sentence $X$, $v^e(X)$ and $v^e(\mathsf{T}(\vDash X^{\urcorner}))$ are the same. If we can do this then we have ways in which $\mathsf{T}$ can behave like a truth predicate. Following ideas that started with Kripke (1975) and Martin and Woodruff (1975), we use what is often called a *truth revision operator* where, if $v$ is a $\mathsf{T}$ valuation, then applying a truth revision operator to $v$ produces another $\mathsf{T}$ valuation

that explicitly incorporates some of the information that was implicit in $v$ itself. We then look for valuations that are unchanged by this process.

**Definition 7.6.3** (Truth Revision Operator) Let $e$ be a valuation extension in $\mathcal{B}$. An *e truth revision operator* on bilattice $\mathcal{B}$ is a mapping $\varphi^e_{\mathcal{B}} : \mathcal{B}^{\mathcal{A}(\mathsf{T})} \rightarrow \mathcal{B}^{\mathcal{A}(\mathsf{T})}$ such that for each $\mathsf{T}$ valuation $v$ in $\mathcal{B}$, $\varphi^e_{\mathcal{B}}(v)$ is the $\mathsf{T}$ valuation in $\mathcal{B}$ such that $\varphi^e_{\mathcal{B}}(v)(\mathsf{T}(\ulcorner X \urcorner)) = v^e(X)$ for every sentence $X$ of $\mathcal{L}$.

A fixed point of truth revision operator $\varphi^e_{\mathcal{B}}$ is a $\mathsf{T}$ valuation $v$ such that $v = \varphi^e_{\mathcal{B}}(v)$ and thus, for every sentence $X$, $v(\mathsf{T}(\ulcorner X \urcorner)) = \varphi^e_{\mathcal{B}}(v)(\mathsf{T}(\ulcorner X \urcorner)) = v^e(X)$. Then with respect to fixed point valuations, the truth predicate $\mathsf{T}$ does have the fundamental property we want a truth predicate to have. Of course liar sentences show this can't happen in a classical framework, but the additional machinery of bilattices gives us exactly what we need. To show fixed points exist in a bilattice, the key thing needed is monotonicity, and then we can make use of Knaster-Tarski. We saw results called monotonicity in Propositions 7.4.6, 7.4.8, and 7.4.10. The following is a general characterization, bringing the $\mathsf{T}$ predicate into things.

**Definition 7.6.4** (Monotonicity Property) We say the *monotonicity property* holds for valuation extension $e$ in bilattice $\mathcal{B}$ provided, for any $\mathsf{T}$ valuations, if $v_1 \leq_k v_2$ in $\mathcal{B}^{\mathcal{A}(\mathsf{T})}$ then $v_1^e \leq_k v_2^e$ in $\mathcal{B}^{\mathcal{L}}$.

**Proposition 7.6.5** *If valuation extension e has the monotonicity property in bilattice $\mathcal{B}$ then the truth revision operator $\varphi^e_{\mathcal{B}}$ is monotonic in the information ordering in $\mathcal{B}^{\mathcal{A}(\mathsf{T})}$, that is, $v_1 \leq_k v_2$ implies $\varphi^e_{\mathcal{B}}(v_1) \leq_k \varphi^e_{\mathcal{B}}(v_2)$.*

***Proof*** Assume $e$ has the monotonicity property, and $v_1 \leq_k v_2$. To show that $\varphi^e_{\mathcal{B}}(v_1) \leq_k \varphi^e_{\mathcal{B}}(v_2)$ we show that for every sentence $X$ we have $\varphi^e_{\mathcal{B}}(v_1)(\mathsf{T}(\ulcorner X \urcorner)) \leq_k \varphi^e_{\mathcal{B}}(v_2)(\mathsf{T}(\ulcorner X \urcorner))$. But by Definition 7.6.3, this is equivalent to $v_1^e(X) \leq_k v_2^e(X)$, and we have this by the monotonicity property, Definition 7.6.4, since we have $v_1 \leq_k v_2$. □

**Corollary 7.6.6** *In bilattice $\mathcal{B}$, if valuation extension e has the monotonicity property then the corresponding truth revision operator $\varphi^e_{\mathcal{B}}$ has fixed points, in particular it has smallest and greatest ones with respect to the $\leq_k$ bilattice ordering.*

This Corollary is immediate by the Knaster-Tarski result, Proposition 7.5.2, though at the moment we don't officially know that there are valuation extensions having the monotonicity property. In fact there are several, as will be shown in Sect. 7.8.

## 7.7 Conflation and Consistency

When looking at $\mathcal{FOUR}$ in Fig. 7.1 certain substructures naturally present themselves. The subset $\{\mathbf{f}, \mathbf{t}\}$ suggests classical logic; the truth operations $\wedge, \vee, \neg$ restricted to $\{\mathbf{f}, \mathbf{t}\}$ have exactly the classical behavior. Similarly the subset $\{\mathbf{f}, \bot, \mathbf{t}\}$ behaves

like the strong Kleene logic $\mathsf{K}_3$. Dually, the subset $\{\mathbf{f}, \top, \mathbf{t}\}$ can be thought of as like Priest's logic of paradox $\mathsf{LP}$. In some sense the lower values in $\mathcal{FOUR}$ represent consistency, while the upper values represent a kind of dual notion. (We will use "lower" and "upper" informally here, corresponding to position in the $\leq_k$ bilattice ordering.) As it turns out, analogs of these structures can be found in bilattices generally provided we have a conflation operator, something that has played little role for us up to this point.

As motivation let $S$ be a set of agents (recall Example 7.4.1), with $\mathcal{P}(S)$ as the collection of all subsets, and with $\subseteq$ as a partial ordering. This is a De Morgan lattice with complementation, $X \mapsto (S - X)$, as involution (Definition 7.3.2). The bilattice product, $\mathcal{P}(S) \odot \mathcal{P}(S)$, is a bilattice with conflation. In it a value is a pair $\langle F, A\rangle$ where $F$ is the set of agents with an opinion *for* some proposition, and $A$ is the set with an opinion *against*. An agent may have an opinion both for and against, in which case $\langle F, A\rangle$ can be thought of as representing an inconsistency. An agent may have no opinion either way, in which case $\langle F, A\rangle$ is an incomplete bilattice value. Or it may be that every agent falls into exactly one of $F$ or $A$, and so we have a kind of exactness of opinions. We incorporate these ideas into bilattices in a structural way. If every agent falls into exactly one of the *for* or *against* categories the corresponding bilattice value is of the form $\langle F, \overline{F}\rangle$, and values like these are their own conflations. If no agent is both *for* and *against*, the bilattice value is of the form $\langle F, A\rangle$ where $F \cap A = \emptyset$, or equivalently where $F \subseteq \overline{A}$, or $A \subseteq \overline{F}$. Using conflation this can be expressed quite simply as $\langle F, A\rangle \leq_k -\langle F, A\rangle$. That some agent is both *for* and *against* corresponds to $-\langle F, A\rangle \leq_k \langle F, A\rangle$ but $-\langle F, A\rangle \neq \langle F, A\rangle$. This leads us to the following.

**Definition 7.7.1** (Consistent, Exact, Anticonsistent) A member $x$ of a bilattice with conflation is *consistent* if $x \leq_k -x$, *anticonsistent* if $-x \leq_k x$, and *exact* if $x = -x$.

This three way division does not always exhaust a bilattice. $\mathcal{SIXTEEN}$, in Fig. 7.3, has a conflation, $\langle \bot, \top\rangle$ and $\langle \top, \bot\rangle$ are conflations of each other but are not comparable in the $\leq_k$ ordering, and so are neither exact, consistent, nor anticonsistent. Such values merit further investigation, but not here. In $\mathcal{FOUR}$ the consistent members correspond to $\mathsf{K}_3$, the anticonsistent members to $\mathsf{LP}$, and the exact members to classical logic. The results that follow say these categories, exhaustive or not, always give us coherent subsystems of bilattices with conflation.

**Proposition 7.7.2** *In a bilattice with conflation we have the following.*

1. *Each of the consistent, anticonsistent, and exact subsets are closed under truth operations $\wedge$, $\vee$, and $\neg$, as well as under infinitary operations $\bigwedge$ and $\bigvee$.*
2. *The consistent values are closed under chain sups with respect to the $\leq_k$ ordering, Definition 7.5.3.*
3. *Dually the anticonsistent values are closed under chain infs.*
4. *Anything below a consistent value, in the $\leq_k$ ordering, is consistent.*

***Proof*** We work in a biattice with conflation.

1. If $x$ and $y$ are consistent, $x \leq_k -x$ and $y \leq_k -y$, then by interlacing, $x \wedge y \leq_k -x \wedge -y$, and by Proposition 7.2.5 $-x \wedge -y = -(x \wedge y)$, so $x \wedge y \leq_k -(x \wedge y)$, and hence $x \wedge y$ is consistent. Using infinitary interlacing a similar proof gives that if every member of $S$ is consistent then $\bigwedge S$ is consistent. The cases with $\vee$ and $\bigvee$ are similar. If $x \leq_k -x$ then $\neg x \leq_k \neg - x$, so $\neg x \leq_k -\neg x$ since negation and conflation commute, so $\neg x$ is consistent. Anticonsistent and exact are similar.
2. Suppose $S$ is a chain of consistent bilattice values. We first show that for any $a, b \in S$ we have $a \leq_k -b$. Since both $a$ and $b$ are in the chain $S$, either $a \leq_k b$ or $b \leq_k a$. Since members of $S$ are consistent, $a \leq_k -a$ and $b \leq -b$. If $a \leq_k b$, since $b \leq_k -b$, then $a \leq_k -b$. If $b \leq_k a$, since $a \leq_k -a$, then $b \leq_k -a$ so $a \leq_k -b$. Then either way, $a \leq_k -b$.
   Let $-S = \{-x \mid x \in S\}$. Now let $a$ and $b$ be arbitrary members of $S$. Then $-b$ is an arbitrary member of $-S$, and by what was just shown, $a \leq_k -b$. Since $-b$ is arbitrary, $a \leq -b$ for every $-b \in -S$, so $a$ is a lower bound for $-S$, and hence $a \leq_k \prod -S$. Since $a$ was arbitrary, $\prod -S$ is an upper bound for $S$, and so $\sum S \leq_k \prod -S$. Then by Proposition 7.2.5, $\sum S \leq_k -\sum S$, and so $\sum S$ is consistent.
3. By an argument dual to the previous one.
4. Suppose $a$ is consistent (so $a \leq_k -a$) and $b \leq_k a$ (and so $-a \leq_k -b$). Then $b \leq_k a \leq_k -a \leq_k -b$, so $b$ is consistent. □

Next, something that will play a role when we come to supervaluations in Sect. 7.8.4.

**Proposition 7.7.3** *In a bilattice with conflation, if $x$ is consistent then $x \leq_k y$ for some exact $y$, and $x = \prod\{y \mid y \text{ is exact and } x \leq_k y\}$.*

***Proof*** A direct proof of the second part is possible but ornate, to say the least. In Sect. 7.3.1 we noted that the bilattice product construction was universal. Then if we prove the result for product bilattices $L \odot L$, where $L$ is any non-distributive De Morgan lattice, we have the result generally.

Assume $\langle x_1, x_2\rangle \in L \odot L$, as in Definition 7.3.1, and is consistent, so $\langle x_1, x_2\rangle \leq_k -\langle x_1, x_2\rangle = \langle \overline{x_2}, \overline{x_1}\rangle$, which unwinds to $x_1 \leq \overline{x_2}$, or equivalently, $x_2 \leq \overline{x_1}$. Let $S = \{\langle y_1, y_2\rangle \mid \langle y_1, y_2\rangle \text{ is exact and } \langle x_1, x_2\rangle \leq_k \langle y_1, y_2\rangle\}$. We show $\langle x_1, x_2\rangle = \prod S$.

Trivially $\langle x_1, x_2\rangle \leq_k \prod S$. Also $\langle x_1, \overline{x_1}\rangle$ is easily seen to be exact and $\langle x_1, x_2\rangle \leq_k \langle x_1, \overline{x_1}\rangle$ because $x_2 \leq \overline{x_1}$, so $\langle x_1, \overline{x_1}\rangle \in S$ (incidentally giving the first part of the Proposition). Similarly $\langle \overline{x_2}, x_2\rangle \in S$. And $\langle x_1, \overline{x_1}\rangle \otimes \langle \overline{x_2}, x_2\rangle = \langle x_1 \sqcap \overline{x_2}, \overline{x_1} \sqcap x_2\rangle = \langle x_1, x_2\rangle$ because $x_1 \leq \overline{x_2}$. Then $\prod S \leq_k \langle x_1, \overline{x_1}\rangle \otimes \langle \overline{x_2}, x_2\rangle = \langle x_1, x_2\rangle$. □

Part 1 of Proposition 7.7.2 says that bilattices with conflation have subsystems that are natural generalizations of $\mathsf{K}_3$, $\mathsf{LP}$, and classical logic. Kripke's work was entirely within the consistent part of $\mathcal{FOUR}$. We now discuss what is needed to recover his results—what conditions keep us within the consistent parts of bilattices. Note that if $\mathcal{B}$ is a bilattice with conflation, so is $\mathcal{B}^{\mathcal{A}(\mathsf{T})}$; and $v \in \mathcal{B}^{\mathcal{A}(\mathsf{T})}$ is consistent if and only if $v(\mathsf{T}(0^{++\cdots+}))$ is consistent in $\mathcal{B}$ for every numeral, and similarly for anticonsistency.

**Definition 7.7.4** Let $\mathcal{B}$ be a bilattice with conflation. Valuation extension *e preserves consistency* if, for every consistent $v \in \mathcal{B}^{\mathcal{A}(\mathsf{T})}$, the $\mathsf{T}$ valuation $\varphi^e_{\mathcal{B}}(v)$ is consistent, where $\varphi^e_{\mathcal{B}}$ is the truth revision operator. Similarly for *preserving anticonsistency*.

**Proposition 7.7.5** *Let $\mathcal{B}$ be a bilattice with conflation. If valuation extension e preserves consistency then the least fixed point of operator $\varphi^e_{\mathcal{B}}$ is consistent in $\mathcal{B}^{\mathcal{A}(\mathcal{T})}$. Similarly if e preserves anticonsistency then the greatest fixed point is anticonsistent.*

***Proof*** We show the set of consistent members of the bilattice $\mathcal{B}^{\mathcal{A}(\mathsf{T})}$ is $\varphi^e_{\mathcal{B}}$ inductive starting at $\bot$ (using the $\leq_k$ ordering), Definition 7.5.3.

The least member $\bot$ of $\mathcal{B}^{\mathcal{A}(\mathsf{T})}$ is consistent because it is below everything, in particular its conflation. If $v$ is a $\mathsf{T}$ valuation that is consistent in $\mathcal{B}^{\mathcal{A}(\mathsf{T})}$ then $\varphi^e_{\mathcal{B}}(v)$ is consistent because we are assuming preservation of consistency. Finally the consistent values are closed under chain sups by part 2 of Proposition 7.7.2.

The result follows by Corollary 7.5.4. And the entire argument dualizes. □

## 7.8 Particular Valuation Extensions

Valuation extensions were introduced in Sect. 7.6, specifically in Definition 7.6.2. Many different ones are plausible and interesting. We look at four of them.

### *7.8.1 Kleene's Strong Three Valued Logic, Generalized*

We use $s$ for the valuation extension map introduced below to suggest *strong*, as in Kleene's strong three valued logic, $\mathsf{K}_3$. We emphasize the Kleene logic since it played a big role in Kripke (1975), but what we do also generalizes Graham Priest's three-valued Logic of Paradox, $\mathsf{LP}$, and the Belnap-Dunn $\mathcal{FOUR}$. It might be helpful to review Example 7.4.9.

**Definition 7.8.1** (Generalized Strong Kleene) Let $\mathcal{B}$ be a bilattice. The strong Kleene valuation extension is the mapping $s : \mathcal{B}^{\mathcal{A}(\mathsf{T})} \to \mathcal{B}^{\mathcal{L}}$ where, for $v \in \mathcal{B}^{\mathcal{A}(\mathsf{T})}$, $v^s \in \mathcal{B}^{\mathcal{L}}$ simply uses the operations of $\mathcal{B}$ from the $\leq_t$ ordering, plus arithmetic facts.

1. For atomic sentences involving $\mathsf{T}$, $v^s(\mathsf{T}(0^{++\cdots+})) = v(\mathsf{T}(0^{++\cdots+}))$.
2. If $X$ is an atomic sentence not involving $\mathsf{T}$, it will be a sentence of arithmetic. It is understood as in the standard model for arithmetic. Specifically, for $a$, $b$, and $c$ being numerals that name numbers $a^\circ$, $b^\circ$, and $c^\circ$ in the standard model:
   a. $v^s(A(a,b,c)) = \mathbf{t}$ in $\mathcal{B}$ if $a^\circ + b^\circ = c^\circ$, and otherwise $v^s(A(a,b,c)) = \mathbf{f}$;
   b. $v^s(M(a,b,c)) = \mathbf{t}$ in $\mathcal{B}$ if $a^\circ \times b^\circ = c^\circ$, and otherwise $v^s(M(a,b,c)) = \mathbf{f}$;
   c. $v^s(a = b) = \mathbf{t}$ if $a^\circ = b^\circ$, and otherwise $v^s(a = b) = \mathbf{f}$.
3. $v^s(X \wedge Y) = v^s(X) \wedge v^s(Y)$

4. $v^s(X \vee Y) = v^s(X) \vee v^s(Y)$
5. $v^s(\neg X) = \neg v^s(X)$
6. $v^s(\forall x F(x)) = \bigwedge\{v^s(F(t)) \mid t \text{ is a numeral}\}$
7. $v^s(\exists x F(x)) = \bigvee\{v^s(F(t)) \mid t \text{ is a numeral}\}$

If language $\mathcal{L}$ is modal, we add the following, for $\mathcal{F} = \langle \mathcal{G}, \mathcal{R} \rangle$ a modal frame.

8. $v^s(\Box X)(\Gamma) = \bigwedge\{v^s(X)(\Delta) \mid \Delta \in \mathcal{G} \text{ and } \Gamma \mathcal{R} \Delta\}$
9. $v^s(\Diamond X)(\Gamma) = \bigvee\{v^s(X)(\Delta) \mid \Delta \in \mathcal{G} \text{ and } \Gamma \mathcal{R} \Delta\}$

This gives us the monotonicity we need, with or without modal operators, Definition 7.6.4. Note that $v$ and $v^s$ agree on atomic sentences $\mathsf{T}(0^{++\cdots+})$.

**Lemma 7.8.2** (Strong Kleene Monotonicity) *If $v_1 \leq_k v_2$ in $\mathcal{B}^{\mathcal{A}(\mathsf{T})}$, then $v_1^s \leq_k v_2^s$ in $\mathcal{B}^{\mathcal{L}}$.*

***Proof*** Assume $v_1 \leq_k v_2$ and show $v_1^s(X) \leq_k v_2^s(X)$ by induction on the complexity of sentence $X$.

At the atomic level we have the result for sentences of the form $\mathsf{T}(0^{++\cdots+})$ because $v_1 \leq_k v_2$. We have it for arithmetic sentences because $v_1^s$ and $v_2^s$ agree on them.

For the induction step, if we have $X \wedge Y$ and the result is known for simpler formulas, then using the interlacing conditions, $v_1^s(X \wedge Y) = v_1^s(X) \wedge v_1^s(Y) \leq_k v_2^s(X) \wedge v_2^s(Y) = v_2^s(X \wedge Y)$. Similarly for $\vee$, while $\neg$ follows using the negation condition that $x \leq_k y$ implies $\neg x \leq_k \neg y$. The result follows in the quantifier cases using infinitary interlacing, and similarly for the modal cases. □

Now by Corollary 7.6.6, a strong Kleene truth revision operator has a smallest and a biggest fixed point in the bilattice $\mathcal{B}^{\mathcal{A}(\mathsf{T})}$ of $\mathsf{T}$ valuations in a bilattice $\mathcal{B}$. We can go further. The first part of what follows covers Kripke's use of Kleene's $\mathsf{K}_3$, and the second part says how Priest's $\mathsf{LP}$ figures in.

**Proposition 7.8.3** *For a bilattice $\mathcal{B}$ with conflation, the least fixed point of the strong Kleene truth revision operator is consistent and the greatest is anticonsistent.*

***Proof*** Using Proposition 7.7.5, for the first part it is enough to show that $\varphi_{\mathcal{B}}^s$ preserves consistency, and for the second part that it preserves anticonsistency.

Suppose $v$ is a $\mathsf{T}$ valuation that is consistent. To show $\varphi_{\mathcal{B}}^s(v)$ is consistent we show that, for every sentence $X$, $\varphi_{\mathcal{B}}^s(v)(\mathsf{T}(\ulcorner X \urcorner))$ is a consistent member of $\mathcal{B}$. Since $v$ is consistent then $v(\mathsf{T}(t))$ is a consistent member of $\mathcal{B}$ for each numeral $t$. The value of $v$ on atomic sentences of arithmetic must be either $\mathbf{t}$ or $\mathbf{f}$, both of which are consistent. Then $v^s$ maps all atomic sentences to consistent values in $\mathcal{B}$, so using Proposition 7.7.2 part 1, $v^s$ maps every sentence of $\mathcal{L}$ to a consistent truth value. In particular, $v^s(X)$ is consistent, but this is $\varphi_{\mathcal{B}}^s(v)(\mathsf{T}(\ulcorner X \urcorner))$.

Anticonsistency is similar and is omitted. □

We now look at what the strong Kleene fixed points can be like for some of the bilattices from Sect. 7.4. Since arithmetic always behaves as it does in the standard model, we can use the Gödel fixed point theorem: for any formula $F(x)$ of language

$\mathcal{L}$ having one free variable, a sentence $X$ can be constructed such that $F(\ulcorner X \urcorner)$ and $X$ are equivalent, in the sense that they always have the same value under $v^s$ for any valuation $v$ in a bilattice.

A *liar sentence* $L$ is equivalent to $\neg\mathsf{T}(\ulcorner L \urcorner)$—a Gödel fixed point of $\neg\mathsf{T}(x)$. In any bilattice $\mathcal{B}$, for any valuation $v$ that is a fixed point of truth revision operator $\varphi^s_{\mathcal{B}}$, it must be the case that $v^s$ assigns to $L$ a bilattice value that is its own negation. For, suppose $v$ were a fixed point valuation. Then $v^s(\mathsf{T}(\ulcorner L \urcorner)) = v(\mathsf{T}(\ulcorner L \urcorner)) = \varphi^s_{\mathcal{B}}(v)(\mathsf{T}(\ulcorner L \urcorner)) = v^s(L)$. But since $L$ and $\neg\mathsf{T}(\ulcorner L \urcorner)$ are equivalent, $v^s(\neg\mathsf{T}(\ulcorner L \urcorner)) = v^s(L)$, so $v^s(L) = \neg v^s(\mathsf{T}(\ulcorner L \urcorner))$. Then $v^s(L) = \neg v^s(L)$. It follows that in no bilattice $\mathcal{B}$ can truth revision operator $\varphi^s_{\mathcal{B}}$ have a fixed point valuation in which the liar sentence $L$ takes on either value **f** or **t**, for neither of these are negations of themselves.

In $\mathcal{FOUR}$, by the Knaster-Tarski theorem the strong Kleene truth revision operator has a smallest and a greatest fixed point under $\leq_k$. The liar sentence $L$ cannot be assigned either **t** or **f** by fixed points, so the values assigned to the liar in $\mathcal{FOUR}$ must be $\bot$ or $\top$. The smallest fixed point is consistent and the biggest is anticonsistent.

More complicated bilattices than $\mathcal{FOUR}$ have technical interest and may have philosophical interest as well. Consider Fig. 7.2 showing the bilattice $\mathcal{NINE}$. In this the smallest and biggest fixed point valuations still give the liar the values $\bot = \langle 0, 0\rangle$ and $\top = \langle 1, 1\rangle$, as they did in $\mathcal{FOUR}$. But in $\mathcal{NINE}$ there is one more value that is its own negation, $\langle \frac{1}{2}, \frac{1}{2}\rangle$, which suggests it is a value that the liar could take on in some fixpoint. In fact this happens, and thus in $\mathcal{NINE}$ there are at least three fixpoint valuations that handle the liar. To show there actually is such a fixed point we can use the Knaster-Tarski Theorem 7.5.2. Here is the construction.

Suppose $L$ and $\neg\mathsf{T}(\ulcorner L \urcorner)$ are equivalent. Let $v$ be the valuation in $\mathcal{NINE}$ such that $v(\mathsf{T}(\ulcorner L \urcorner)) = \langle \frac{1}{2}, \frac{1}{2}\rangle$, and for every sentence $X$ other than $L$, $v(\mathsf{T}(\ulcorner X \urcorner)) = \langle 0, 0\rangle$, that is, the value is $\bot$ of $\mathcal{NINE}$. To keep notation uncluttered, let $\varphi = \varphi^s_{\mathcal{NINE}}$ be the strong Kleene truth revision operator for $\mathcal{NINE}$. Then $\varphi(v)(\mathsf{T}(\ulcorner L \urcorner)) = v^s(L) = v^s(\neg\mathsf{T}(\ulcorner L \urcorner)) = \neg v^s(\mathsf{T}(\ulcorner L \urcorner)) = \neg\langle \frac{1}{2}, \frac{1}{2}\rangle = \langle \frac{1}{2}, \frac{1}{2}\rangle = v(\mathsf{T}(\ulcorner L \urcorner))$, so $v(\mathsf{T}(\ulcorner L \urcorner)) = \varphi(v)(\mathsf{T}(\ulcorner L \urcorner))$. On all other sentences $X$, $v(\mathsf{T}(\ulcorner X \urcorner)) \leq_k \varphi(v)(\mathsf{T}(\ulcorner X \urcorner))$ since $v(\mathsf{T}(\ulcorner X \urcorner)) = \bot$. Then $v \leq_k \varphi(v)$. Now by Knaster-Tarski 7.5.2, part 2, there is a fixed point valuation $v_f$ of $\varphi$ above $v$ in the $\leq_k$ ordering. It is easy to check that every valuation in the Ordinal Sequence approximating to $v_f$ will assign $\langle \frac{1}{2}, \frac{1}{2}\rangle$ to $\ulcorner L \urcorner$, hence this is what $v_f$ assigns.

In Example 7.4.4 we briefly discussed a Fuzzy Bilattice, the bilattice product of the closed unit interval with itself. Taking this as the space of truth values gives us a kind of 'degree of confidence' structure. The truth revision operator in the bilattice of valuations in this structure has fixpoints. The construction just presented for $\mathcal{NINE}$ can easily be transferred. Every fuzzy truth value of the form $\langle a, a\rangle$ is a value the liar sentence can take on in some strong Kleene fixpoint using the fuzzy bilattice.

### 7.8.2 *Kleene's Weak Three Valued Logic, Generalized*

A three valued logic that originated in Bochvar (1938) was introduced independently in Kleene (1950) with a different motivation. Commonly called *Kleene's weak three valued logic*, it was one of the logics employed in Kripke (1975), and the only logic used in Martin and Woodruff (1975). It behaves classically on classical truth values, but simply assigns $\bot$ to any formula under a valuation whenever any atomic subformula is assigned $\bot$ by that valuation. It can be reformulated in a way that has an intuition extending beyond the three valued setting. This was foreshadowed in Fitting (1994), and given its present form in Fitting (2006) under the name "cut down operations". For recent work on cut down operations, see Ferguson (2015).

As motivation we turn once again to bilattices arising from the opinions of sets of agents, continuing the discussions from Sect. 7.7 and Example 7.4.1. Recall that the setting is a product bilattice $\mathcal{P}(S) \odot \mathcal{P}(S)$, where $\mathcal{P}$ is the power set operation, $S$ is a set of agents, and the ordering on $\mathcal{P}(S)$ is $\subseteq$. In this bilattice a truth value is $\langle F, A\rangle$ with the $F$ values informally thought of as those agents *for* and the $A$ values as those *against* some proposition. But now in evaluating, say, $X \wedge Y$ suppose we only want to consider the opinions of agents on $Y$ who have actually expressed an opinion on $X$, and similarly the other way around. That is, we want to "cut down" the sets of agents to the completely active ones. Suppose the truth value assigned to $X$ is $\langle F, A\rangle$; then the active agents on $X$ are $F \cup A$. The consensus of $\langle F \cup A, F \cup A\rangle$ with the truth value assigned to $Y$ would give us the kind of restriction we need. But also, $\langle F \cup A, F \cup A\rangle = \langle F, A\rangle \oplus \langle A, F\rangle = \langle F, A\rangle \oplus \neg\langle F, A\rangle$. This leads us to the following, where the superscript $w$ is meant to suggest *weak*, as in Kleene's *weak* logic.

**Definition 7.8.4** (Cut Down Operations) Let $\mathcal{B}$ be a bilattice with $x$ and $y$ as members, and $S$ as a set of members.

1. $\|x\| = x \oplus \neg x$
2. $x \wedge^w y = (x \wedge y) \otimes \|x\| \otimes \|y\|$
3. $x \vee^w y = (x \vee y) \otimes \|x\| \otimes \|y\|$
4. $\bigwedge^w S = (\bigwedge S) \otimes \prod\{\|x\| \mid x \in S\}$
5. $\bigvee^w S = (\bigvee S) \otimes \prod\{\|x\| \mid x \in S\}$

We did not define a weak negation. If we had, it would have been $\neg^w x = \neg x \otimes \|x\|$, but this is easily seen to be the same as $\neg x$.

The cut down operations restricted to the consistent part of $\mathcal{FOUR}$, $\{\mathbf{f}, \bot, \mathbf{t}\}$, are exactly the weak Kleene or Bochvar operations. Indeed, the motivation for the cut down operations suggests this should be so since this is the one agent case, and if the agent has no opinion about something, that lack of opinion propagates through any application of the weak operations. Of course our logic operations are actually on the whole of $\mathcal{FOUR}$, and the extension follows the familiar pattern. Throughout $\mathcal{FOUR}$, $\wedge^w$ and $\vee^w$ behave the same as $\wedge$ and $\vee$, except that any conjunction or disjunction with a component evaluating to $\bot$ itself evaluates to $\bot$.

**Proposition 7.8.5** *In any bilattice with conflation the set of consistent values is closed under the truth operations* $\wedge^w$, $\vee^w$, *and under the infinitary operations* $\bigwedge^w$ *and* $\bigvee^w$.

***Proof*** Suppose both $x$ and $y$ are consistent. By Proposition 7.7.1, $x \wedge y$ is consistent and so $x \wedge y \leq_k -(x \wedge y)$. Then,

$$\begin{aligned} x \wedge^w y &= (x \wedge y) \otimes \|x\| \otimes \|y\| \\ &\leq_k (x \wedge y) \leq_k -(x \wedge y) \\ &\leq_k -(x \wedge y) \oplus -\|x\| \oplus -\|y\| \\ &= -[(x \wedge y) \otimes \|x\| \otimes \|y\|] \\ &= -(x \wedge^w y) \end{aligned}$$

so $x \wedge^w y$ is consistent. The other cases are similar. □

Note that, unlike with our strong Kleene generalization in Proposition 7.7.2, we did not claim closure of the anticonsistent or exact part under our weak logic connectives.

**Example 7.8.6** In the bilattice $\mathcal{SIXTEEN}$ from Fig. 7.3 neither the anticonsistent nor the exact values are closed under $\wedge^w$. Recall, this is the bilattice product of $\mathcal{FOUR}$ from Fig. 7.1 with itself, using the $\leq_t$ lattice ordering for $\mathcal{FOUR}$. With this ordering, $\mathcal{FOUR}$ is a De Morgan lattice with $\overline{\top} = \top$, $\overline{\bot} = \bot$, $\bar{\mathbf{t}} = \mathbf{f}$, and $\bar{\mathbf{f}} = \mathbf{t}$.

In $\mathcal{SIXTEEN}$ $-\langle \bot, \bot \rangle = \langle \overline{\bot}, \overline{\bot} \rangle = \langle \bot, \bot \rangle$ and similarly $-\langle \top, \top \rangle = \langle \top, \top \rangle$, so both $\langle \bot, \bot \rangle$ and $\langle \top, \top \rangle$ are exact, and hence are both anticonsistent and consistent as well. Now $\langle \bot, \bot \rangle \wedge^w \langle \top, \top \rangle = (\langle \bot, \bot \rangle \wedge \langle \top, \top \rangle) \otimes \|\langle \bot, \bot \rangle\| \otimes \|\langle \top, \top \rangle\| = \langle \mathbf{f}, \mathbf{t} \rangle \otimes \langle \bot, \bot \rangle \otimes \langle \top, \top \rangle = \langle \mathbf{f}, \mathbf{f} \rangle$. But $-\langle \mathbf{f}, \mathbf{f} \rangle = \langle \bar{\mathbf{f}}, \bar{\mathbf{f}} \rangle = \langle \mathbf{t}, \mathbf{t} \rangle$ so $\langle \mathbf{f}, \mathbf{f} \rangle$ is not exact, and it is not anticonsistent since $\langle \mathbf{t}, \mathbf{t} \rangle \not\leq_k \langle \mathbf{f}, \mathbf{f} \rangle$. In accord with Proposition 7.8.5, it should be consistent and, in fact, we do have $\langle \mathbf{f}, \mathbf{f} \rangle \leq_k \langle \mathbf{t}, \mathbf{t} \rangle$.

More generally, suppose $L \odot L$ is a bilattice product, where $L$ is a non-distributive De Morgan lattice. It can be shown that, in $L \odot L$, the exact values are closed under the weak logic connectives $\wedge^w$ and $\vee^w$ if and only if, in $L$, $x \vee \overline{x} = 1$ and $x \wedge \overline{x} = 0$ for all $x$. This also allows us to conclude that the exact values of $\mathcal{SIXTEEN}$ are not closed under $\wedge^w$ and $\vee^w$.

We now define the obvious weak Kleene valuation extension, an instance of Definition 7.6.2. It is well behaved on the consistent part of a bilattice with conflation and at least makes sense outside that range, but its behavior there has not been investigated. The definition has been dualized in Szmuc (2018), producing what are called *track down* operations. These are well behaved on the anticonsistent part, and so produce a weak generalization of Priest's LP.

**Definition 7.8.7** (Generalized Weak Kleene) Let $\mathcal{B}$ be a bilattice. The weak Kleene valuation extension is the map $w : \mathcal{B}^{\mathcal{A}(\mathsf{T})} \to \mathcal{B}^{\mathcal{L}}$ where, for $v \in \mathcal{B}^{\mathcal{A}(\mathsf{T})}$, $v^w \in \mathcal{B}^{\mathcal{L}}$ uses the cut down operations on $\mathcal{B}$.

1. For atomic sentences involving $\mathsf{T}$, $v^w(\mathsf{T}(0^{++\cdots+})) = v^s(\mathsf{T}(0^{++\cdots+})) = v(\mathsf{T}(0^{++\cdots+}))$.

2. For numerals $a$, $b$, and $c$:
   a. $v^w(A(a,b,c)) = v^s(A(a,b,c))$;
   b. $v^w(M(a,b,c)) = v^s(M(a,b,c))$;
   c. $v^w(a = b) = v^s(a = b)$.
3. $v^w(X \wedge Y) = v^w(X) \wedge^w v^w(Y)$
4. $v^w(X \vee Y) = v^w(X) \vee^w v^w(Y)$
5. $v^w(\neg X) = \neg v^w(X)$
6. $v^w(\forall x F(x)) = \bigwedge^w \{v^w(F(t)) \mid t \text{ is a numeral}\}$
7. $v^w(\exists x F(x)) = \bigvee^w \{v^w(F(t)) \mid t \text{ is a numeral}\}$

If language $\mathcal{L}$ is modal, we add the following, where $\mathcal{F} = \langle \mathcal{G}, \mathcal{R} \rangle$ is a modal frame.

8. $v^w(\Box X)(\Gamma) = \bigwedge^w \{v^w(X)(\Delta) \mid \Delta \in \mathcal{G} \text{ and } \Gamma \mathcal{R} \Delta\}$
9. $v^w(\Diamond X)(\Gamma) = \bigvee^w \{v^w(X)(\Delta) \mid \Delta \in \mathcal{G} \text{ and } \Gamma \mathcal{R} \Delta\}$

As with strong Kleene, $v$ and $v^w$ agree on atomic sentences $\mathsf{T}(0^{++\cdots+})$. We now have the Monotonicity Property, Definition 7.6.4 by the following.

**Lemma 7.8.8** (Weak Kleene Monotonicity) *If* $v_1 \leq_k v_2$ *in* $\mathcal{B}^{\mathcal{A}(\mathsf{T})}$, *then* $v_1^w \leq_k v_2^w$ *in* $\mathcal{B}^{\mathcal{L}}$.

***Proof*** Assume $v_1 \leq_k v_2$. We show $v_1^w(X) \leq_k v_2^w(X)$ by induction on the complexity of sentence $X$. We only sketch one case and leave the rest to the reader. Suppose $v_1^w(X) \leq_k v_2^w(X)$ and $v_1^w(Y) \leq_k v_2^w(Y)$. We verify that $v_1^w(X \wedge Y) \leq_k v_2^w(X \wedge Y)$. In the following the inequality step is by the induction hypothesis, the lattice properties of $\oplus$ and $\otimes$, the behavior of $\neg$ with respect to $\leq_k$, and the interlacing conditions.

$$\begin{aligned}
v_1^w(X \wedge Y) &= v_1^w(X) \wedge^w v_1^w(Y) \\
&= [v_1^w(X) \wedge v_1^w(Y)] \otimes \|v_1^w(X)\| \otimes \|v_1^w(Y)\| \\
&= [v_1^w(X) \wedge v_1^w(Y)] \otimes [v_1^w(X) \oplus \neg v_1^w(X)] \otimes [v_1^w(Y) \oplus \neg v_1^w(Y)] \\
&\leq_k [v_2^w(X) \wedge v_2^w(Y)] \otimes [v_2^w(X) \oplus \neg v_2^w(X)] \otimes [v_2^w(Y) \oplus \neg v_2^w(Y)] \\
&= v_2^w(X) \wedge v_2^w(Y)
\end{aligned}$$

□

Now we have a weak Kleene truth revision operator $\varphi_{\mathcal{B}}^w$, from Definition 7.6.3. By Proposition 7.6.5, $\varphi_{\mathcal{B}}^w$ is monotonic, and hence has least and greatest fixed points. Using Propositions 7.8.5 and 7.8.3, the least fixed point of $\varphi_{\mathcal{B}}^w$ is consistent in a bilattice with conflation. Unlike with strong Kleene, however, the anticonsistent members of a bilattice are not always closed under the extended weak Kleene operations, so we have no general conclusions concerning the greatest fixed point.

Track down operations were mentioned earlier in this section (Szmuc, 2018). Since their general behavior is dual to cut down operations, it is likely that they will yield a version of a fixed point truth theory generalizing one based on a weak version of Priest's LP. This does not seem to have been investigated yet.

**Fig. 7.4** Four-valued tables for $\wedge^a$ and $\vee^a$

| $\wedge^a$ | $\top$ | **t** | **f** | $\bot$ |
|---|---|---|---|---|
| $\top$ | $\top$ | $\top$ | **f** | **f** |
| **t** | $\top$ | **t** | **f** | $\bot$ |
| **f** | **f** | **f** | **f** | **f** |
| $\bot$ | $\bot$ | $\bot$ | $\bot$ | $\bot$ |

| $\vee^a$ | $\top$ | **t** | **f** | $\bot$ |
|---|---|---|---|---|
| $\top$ | $\top$ | **t** | $\top$ | **t** |
| **t** | **t** | **t** | **t** | **t** |
| **f** | $\top$ | **t** | **f** | $\bot$ |
| $\bot$ | $\bot$ | $\bot$ | $\bot$ | $\bot$ |

### *7.8.3 Asymmetric Logics*

There is a natural propositional three valued asymmetric logic that partially generalizes to bilattices. It seems to have had minimal discussion in the literature on theories of truth, though Kripke and others have talked about it informally, and tableau rules were considered in Fitting (1994). It is, in fact, common in programming languages, where it is generally connected with *lazy evaluation*. Suppose we have $X \wedge Y$, where the values of $X$ and $Y$ are to be determined by calls on procedures which may or may not terminate, and so may or may not return a value. Lazy evaluation calls on the $X$ routine first and if that returns **f** then the $Y$ routine is not called, but instead the overall program concludes the conjunction is false. If the $X$ routine returns **t**, only then is a further call on the $Y$ routine made. Either the $X$ or the $Y$ routine may not terminate, but because lazy evaluation works from left to right there is an asymmetric difference between the two. If $X$ does not terminate there is no final value for the conjunction, but if $X$ returns **f**, even if the $Y$ routine does not terminate, the conjunction has value **f**. It is a logic in which commutativity does not hold.

It is possible to define natural generalizations of propositional connectives for an asymmetric logic using the cut down ideas from Sect. 7.8.2. Recall our earlier discussion involving a group of agents. In these terms, for the conjunction $X \wedge Y$, asymmetry would have us only considering the opinions of experts on $Y$ who have said something about $X$, but not the other way around. This idea gives us the following, where the $a$ is for *asymmetric*. (Again negation does not need any special treatment.)

**Definition 7.8.9** (Asymmetric Operations) Let $x, y \in \mathcal{B}$ where $\mathcal{B}$ is a bilattice. Using Definition 7.8.4 part 1:

1. $x \wedge^a y = (x \wedge y) \otimes \|x\|$
2. $x \vee^a y = (x \vee y) \otimes \|x\|$

Figure 7.4 gives tables for $\wedge^a$ and $\vee^a$ when the bilattice is $\mathcal{FOUR}$. Curiously, restricted to the consistent values $\{\mathbf{f}, \bot, \mathbf{t}\}$ the distributive laws hold, something that programmers make use of all the time. But with all four values present distributivity fails. For instance, $A \wedge^a (B \vee^a C)$ and $(A \wedge^a C) \vee^a (B \wedge^a C)$ differ if $A$ has value $\top$, $B$ has value $\bot$, and $C$ is either $\top$ or **t**.

Quantifiers are a problem. Since our domains are, by convention, the natural numbers and we want a left/right bias, we might think of $\forall x F(x)$ as if it were $F(0) \wedge^a F(1) \wedge^a F(2) \wedge^a \ldots$. But numbers code sentences, so the simple

0, 1, 2, ... order of evaluation might not be what we want at all. To allow full flexibility we could think of quantification as infinite conjunction and disjunction with an order specified. Formally, remove quantifiers from $\mathcal{L}$ and introduce infinitary connectives, $\bigwedge_{i\in\omega} A_i$ and $\bigvee_{i\in\omega} A_i$, where the indexing specifies an ordering. We could then think of the infinitary conjunction as being interpreted asymmetrically as $A_1 \wedge^a (A_2 \otimes \|A_1\|) \wedge^a (A_3 \otimes \|A_1\| \otimes \|A_2\|) \wedge^a \ldots$. Similarly for disjunction. More formally, for a sequence $x_1, x_2, x_3, \ldots \in \mathcal{B}$ where $\mathcal{B}$ is a bilattice, we could introduce the infinitary operations $\bigwedge^a_{i\in\omega} x_i = \bigwedge_{i\in\omega}\{x_i \otimes \prod\{\|x_j\| \mid j < i\}\}$ and $\bigvee^a_{i\in\omega} x_i = \bigvee_{i\in\omega}\{x_i \otimes \prod\{\|x_j\| \mid j < i\}\}$. (If $i = 1$, $\prod\emptyset$ occurs. This is $\top$, and $x_1 \otimes \top = x_1$.)

The asymmetric propositional operations from Definition 7.8.9 combined with the infinitary asymmetric replacements for quantification as just described work well up to a point. It is easy to show we have monotonicity results similar to Lemmas 7.8.2 and 7.8.8. But we can *not* define a monotone truth revision operator, Definition 7.6.3, for the elementary reason that infinitary conjunctions and disjunctions do not have Gödel numbers, and so a truth revision definition can't get off the ground.

*Finite* asymmetric conjunctions and disjunctions have some applications in natural languages. One might plausibly add them to the strong or weak Kleene systems, getting more expressive languages still having the same general fixed point behavior. One might allow infinite conjunctions whose conjuncts are given by a recursive function. Gödel numbering could perhaps then be reintroduced. We leave further investigation of this to others.

### *7.8.4 Supervaluations, Generalized*

The least familiar three-valued logic used in Kripke (1975) is based on *supervaluations*, introduced in Van Fraassen (1966). Neither of Kleene's weak or strong three valued logics validate any classical tautologies since, for both logics, if all atomic formulas are assigned $\bot$ so is every formula. Supervaluations are specially crafted to validate tautologies, but the price paid is that supervaluations are not truth functional. We do not calculate via truth tables, but rather by a more complex method.

Kleene's strong logic, $\mathsf{K}_3$, is three valued, but it extends naturally to all of the Belnap-Dunn four valued logic $\mathcal{FOUR}$. Moreover, every bilattice with conflation has a consistent substructure that generalizes $\mathsf{K}_3$, and this always is a natual part of the entire of the bilattice. Kleene's weak three valued logic has generalizations that are most well-behaved on the consistent parts of bilattices with conflation, but the generalizations make sense for the entirety of bilattices. But supervaluation, in its original sense, is inherently a three valued logic. It lives on the consistent part of $\mathcal{FOUR}$, and makes no sense at all on the whole of it. There is a natural generalization of supervaluation to an arbitrary bilattice with conflation, but again only to the consistent part. Since we would like to have a complete lattice so that the Knaster-Tarski theorem can be applied, we will artificially extend our supervaluation generalization beyond the consistent part of a bilattice, to the entire structure. This

does simplify the mathematics, but we are actually interested in behavior only on consistent values.

Here is the original version of supervaluations from Van Fraassen (1966), as used by Kripke. Suppose we have a $\mathsf{K}_3$ valuation $v$ mapping atomic sentences to $\{\mathbf{f}, \bot, \mathbf{t}\}$. That valuation can be 'raised' to a purely classical one by changing it on every atom that maps to $\bot$ so that instead the valuation maps the atom to one of the classical values, $\mathbf{f}$ or $\mathbf{t}$. Call such a valuation a *totally defined extension* of $v$. For any given sentence $X$, the *supervaluation value* assigned to $X$ by $v$ is $\mathbf{t}$ if every totally defined extension of $v$ maps $X$ to $\mathbf{t}$, the value is $\mathbf{f}$ if every totally defined extension of $v$ maps $X$ to $\mathbf{f}$, and the value is $\bot$ otherwise, that is, when totally defined extensions disagree on $X$. To use bilattice terminology, we are working with valuations in the consistent part of $\mathcal{FOUR}$, and totally defined extensions map just to the exact part of $\mathcal{FOUR}$. Notice that if $v$ is consistent in the bilattice of valuations in $\mathcal{FOUR}$ and $w$ is a totally defined extension of $v$ then $w$ is exact and $v \leq_k w$. And the idea of valuations agreeing or disagreeing can be seen as an application of the bilattice consensus operation, $\otimes$. This should help motivate our generalization, which originated in Fitting (2006). Recall that outside the consistent part we will simply assign an arbitrary value, with $\top$ being the appropriate choice.

**Definition 7.8.10** (Generalized Supervaluations) Let $\mathcal{B}$ be a bilattice. The *supervaluation extension* is the map $sv : \mathcal{B}^{\mathcal{A}(\mathsf{T})} \to \mathcal{B}^{\mathcal{L}}$ where, for $v \in \mathcal{B}^{\mathcal{A}(\mathsf{T})}$, $v^{sv} \in \mathcal{B}^{\mathcal{L}}$ (using *sv* for *supervaluation*) is defined as follows.

$$v^{sv} = \begin{cases} \prod\{w^s \mid w \in \mathcal{B}^{\mathcal{A}(\mathsf{T})} \text{ is exact and } v \leq_k w\} & \text{if } v \text{ is consistent} \\ \top & \text{if } v \text{ is not consistent} \end{cases}$$

In the non-consistent case, $\top$ is that of $\mathcal{B}^{\mathcal{L}}$, the function on $\mathcal{L}$ that is identically the $\top$ element of $\mathcal{B}$. In the consistent case this definition builds on top of the strong Kleene valuation extension. Recall that $v$ is consistent in $\mathcal{B}^{\mathcal{A}(\mathsf{T})}$ if $v(\mathsf{T}(0^{++\cdots+}))$ is consistent in $\mathcal{B}$ for every numeral $0^{++\cdots+}$. In the consistent case the infinitary consensus operation is applied to a non-empty set by Proposition 7.7.3. We are not really interested in valuations that are not consistent—this case is here for technical reasons. But in the consistent case the output value, $v^{sv}$, will be consistent because, more generally, if $S$ is any non-empty set of exact bilattice members, $\prod S$ is consistent. Very simply, $S$ is non-empty, so let $s \in S$. Since $s$ is exact it is consistent and since $\prod S \leq_k s$, $\prod S$ is consistent by Proposition 7.7.2 part 4. Also, in the consistent case $v^{sv}$ assigns to arithmetic sentences the values they have in the standard model, because all strong Kleene extensions do this, and we have Proposition 7.7.3.

**Lemma 7.8.11** (Supervaluation Monotonicity) *If* $v_1 \leq_k v_2$ *in* $\mathcal{B}^{\mathcal{A}(\mathsf{T})}$, *then* $v_1^{sv} \leq_k v_2^{sv}$ *in* $\mathcal{B}^{\mathcal{L}}$.

***Proof*** Assume $v_1 \leq_k v_2$. If $v_2$ is not consistent, $v_2^{sv}(X) = \top$ for every $X$, so trivially $v_1^{sv}(X) \leq_k v_2^{sv}(X)$ for all sentences $X$. If $v_2$ is consistent so is $v_1$, by Proposition 7.7.2

part 4, so the first case of Definition 7.8.10 is used for both $v_1^{sv}$ and $v_2^{sv}$. If $w$ is an exact valuation and $v_2 \leq_k w$, then also $v_1 \leq_k w$, so

$$\{w \in \mathcal{B}^{\mathcal{A}(\mathsf{T})} \mid w \text{ is exact and } v_2 \leq_k w\} \subseteq \{w \in \mathcal{B}^{\mathcal{A}(\mathsf{T})} \mid w \text{ is exact and } v_1 \leq_k w\}$$

and so

$$\begin{aligned} v_1^{sv}(X) &= \prod\{w^s(X) \mid w \in \mathcal{B}^{\mathcal{A}(\mathsf{T})} \text{ and } w \text{ is exact and } v_1 \leq_k w\} \\ &\leq_k \prod\{w^s(X) \mid w \in \mathcal{B}^{\mathcal{A}(\mathsf{T})} \text{ and } w \text{ is exact and } v_2 \leq_k w\} \\ &= v_2^{sv}(X). \end{aligned}$$

□

The supervaluation truth revision operator using bilattice $\mathcal{B}$ (Definition 7.6.3) is denoted $\varphi_{\mathcal{B}}^{sv}$. It is monotone by Proposition 7.6.5, and has a smallest and a greatest fixed point under the $\leq_k$ ordering, by Corollary 7.6.6. The greatest fixed point of $\varphi_{\mathcal{B}}^{sv}$ is artificial, but the least fixed is consistent in a bilattice with conflation. In the consistent part of $\mathcal{FOUR}$ this treatment coincides with supervaluations as in Kripke, but the version here applies more generally, including to modal examples.

Supervaluations, in their original three-valued version, map every tautology to **t**. Things are not so simple for our generalization. Use the bilattice $\mathcal{NINE}$ from Example 7.4.2, let $P$ be atomic, and consider the tautology $P \vee \neg P$ under the valuation $v$ that maps $P$ to $\langle 0, \frac{1}{2}\rangle$. The exact values above $\langle 0, \frac{1}{2}\rangle$ are $\langle 0, 1\rangle$ and $\langle \frac{1}{2}, \frac{1}{2}\rangle$. Let $w_1$ be an exact valuation mapping $P$ to the first of these, and $w_2$ be one mapping $P$ to the second. Then $w_1^s(P \vee \neg P) = \langle 0, 1\rangle \vee \neg\langle 0, 1\rangle = \langle 0, 1\rangle \vee \langle 1, 0\rangle = \langle 1, 0\rangle$ and $w_2^s(P \vee \neg P) = \langle \frac{1}{2}, \frac{1}{2}\rangle \vee \neg\langle \frac{1}{2}, \frac{1}{2}\rangle = \langle \frac{1}{2}, \frac{1}{2}\rangle$. But then $(w_1^s \otimes w_2^s)(P \vee \neg P) = \langle 1, 0\rangle \otimes \langle \frac{1}{2}, \frac{1}{2}\rangle = \langle \frac{1}{2}, 0\rangle$, which is not even exact. Further study of generalized supervaluations may be warranted.

As pointed out by a referee, there is a supervaluation dual called subvaluation—see Cobreros (2013) and Teijeiro (2020). It has been discussed for its applications to vagueness, which also supplied early motivation for supervaluations. It is likely that, when examined in the bilattice context, it will have behavior dual to that of supervaluations. So far there seems to have been no work that has been done on this.

## 7.9 Specialized Fixed Points

We have concentrated on least and greatest fixed points of truth revision operators, but there are other interesting varieties. Kripke (1975) introduced *intrinsic* fixed points, which also independently appeared in Manna and Shamir (1977) in a computer science context. *Maximal* fixed points were central in Martin and Woodruff (1975). *Alternating* fixed points come from Yablo (1985). There are probably others out there but these, at least, generalize to the bilattice family. Both Kripke (1975) and Martin

and Woodruff (1975) worked with three valued logics. The counterpart to that in our approach is the *consistent* part of a bilattice, and a consistency restriction is noted in several parts of this section.

**Terminology Convention 7.9.1** To keep clutter down we announce that throughout this entire section $\mathcal{B}$ is a bilattice with conflation. Also, throughout, we always assume that $e$ is a valuation extension that has the monotonicity property (Definition 7.6.4) and preserves consistency (Definition 7.7.4). This includes generalized strong Kleene, generalized weak Kleene, and generalized supervaluations. We assume $\varphi^e_{\mathcal{B}}$ is the corresponding truth revision operator. Finally $\mathcal{L}$ is our first-order arithmetic language, extended with $\mathsf{T}$, and which may or may not contain modal operators.

### 7.9.1 Consistent Maximal Fixed Points

What we call consistent maximal fixed points are interesting in their own right, and play an important role with respect to intrinsic fixed points, which are discussed in Sect. 7.9.2.

**Definition 7.9.2** (Consistent Maximal Fixed Point) A fixed point $v$ of $\varphi^e_{\mathcal{B}}$ is *consistent maximal* if it is consistent, and there is no other consistent fixed point $w$ with $v \leq_k w$.

The main thing we need is that there actually are consistent maximal fixed points. We recall Definition 7.7.4.

**Proposition 7.9.3** *If $v$ is a consistent $\mathsf{T}$ valuation for which $v \leq_k \varphi^e_{\mathcal{B}}(v)$ (a post-fixed point), then there is a consistent maximal fixed point of $\varphi^e_{\mathcal{B}}$ above $v$, and so consistent maximal fixed points exist.*

***Proof*** Let $\mathcal{F} = \{w \in \mathcal{B}^{\mathcal{A}(\mathsf{T})} \mid w \text{ is consistent and } w \leq_k \varphi^e_{\mathcal{B}}(w)\}$. We show every member of $\mathcal{F}$ is below a consistent maximal fixed point, giving us the first part of the Proposition. The second part follows since $\bot$ is a post-fixed point.

Every chain in $\mathcal{F}$ has an upper bound in $\mathcal{F}$; in fact it has a least upper bound. Here is the argument. Let $C$ be a chain in $\mathcal{F}$. Then $\bigsqcup C$ is consistent by Proposition 7.7.2 part 2. Suppose $w \in C$; then $w \leq_k \bigsqcup C$, so by monotonicity, $\varphi^e_{\mathcal{B}}(w) \leq_k \varphi^e_{\mathcal{B}}(\bigsqcup C)$. Since $w \in C \subseteq \mathcal{F}$, $w \leq_k \varphi^e_{\mathcal{B}}(w)$, so it follows that $w \leq_k \varphi^e_{\mathcal{B}}(\bigsqcup C)$ for every $w \in C$, and consequently $\bigsqcup C \leq_k \varphi^e_{\mathcal{B}}(\bigsqcup C)$. Then $\bigsqcup C \in \mathcal{F}$, so $C$ has an upper bound in $\mathcal{F}$.

The ordering $\leq_k$, restricted to $\mathcal{F}$, is still a partial order, so $\mathcal{F}$ is a partially ordered set in which each chain has an upper bound. By Zorn's Lemma each member of $\mathcal{F}$ extends to a maximal member of $\mathcal{F}$. Finally we show that all consistent fixed points of $\varphi^e_{\mathcal{B}}$ are in $\mathcal{F}$, and that maximal members of $\mathcal{F}$ are themselves consistent fixed points. Then every maximal member of $\mathcal{F}$ will be a consistent maximal fixed point.

If $a$ is a consistent fixed point of $\varphi^e_{\mathcal{B}}$ then, trivially, $a \leq_k \varphi^e_{\mathcal{B}}(a)$, so $a \in \mathcal{F}$. Suppose $m$ is any maximal member of $\mathcal{F}$. By definition it is consistent. Also $m \leq_k \varphi^e_{\mathcal{B}}(m)$ so by monotonicity, $\varphi^e_{\mathcal{B}}(m) \leq_k \varphi^e_{\mathcal{B}}(\varphi^e_{\mathcal{B}}(m))$, and by the preservation of consistency property, $\varphi^e_{\mathcal{B}}(m)$ is consistent. Then $\varphi^e_{\mathcal{B}}(m) \in \mathcal{F}$. Since $m$ is maximal in $\mathcal{F}$ and $m \leq_k \varphi^e_{\mathcal{B}}(m)$, it follows that $m = \varphi^e_{\mathcal{B}}(m)$, so $m$ is a fixed point. □

### 7.9.2 *Intrinsic Fixed Points*

Kripke singled out an interesting subclass of fixed points for particular attention, the *intrinsic* ones, which are those compatible with every fixed point. An equivalent characterization also appeared independently in Manna and Shamir (1977). Here is a bilattice generalization that agrees with these on the consistent part of $\mathcal{FOUR}$.

**Definition 7.9.4** (Intrinsic Fixed Point) A fixed point $v$ of $\varphi^e_{\mathcal{B}}$ is *intrinsic* if $v \oplus w$ is consistent, for every consistent fixed point $w$ of $\varphi^e_{\mathcal{B}}$.

There are intrinsic fixed points, and we have already seen one.

**Proposition 7.9.5** *The smallest fixed point of $\varphi^e_{\mathcal{B}}$ is intrinsic, and every intrinsic fixed point is consistent.*

***Proof*** Let $v$ be the smallest fixed point with respect to $\leq_k$. If $w$ is any consistent fixed point then $v \leq_k w$ so $v \oplus w = w$ which is consistent. Hence $v$ is intrinsic.

Let $w$ be any intrinsic fixed point. Since the smallest fixed point $v$ is consistent, $v \oplus w$ must be consistent because $w$ is intrinsic. But $v \oplus w$ is $w$, hence $w$ is consistent. □

There is a direct connection between intrinsic fixed points and the consistent maximal ones of Sect. 7.9.1.

**Proposition 7.9.6** *A fixed point $v$ of $\varphi^e_{\mathcal{B}}$ is intrinsic if and only if $v \leq_k m$ for every consistent maximal fixed point if and only if $v \leq_k \prod \mathcal{M}$ where $\mathcal{M}$ is the set of all consistent maximal fixed points.*

***Proof*** The if direction and the only if direction have different arguments.

1. Let $v$ be an intrinsic fixed point, and $m$ be a consistent maximal fixed point; we show $v \leq_k m$. Since $v$ is intrinsic, $v \oplus m$ is consistent. Since $v \leq_k v \oplus m$, by monotonicity $\varphi^e_{\mathcal{B}}(v) \leq_k \varphi^e_{\mathcal{B}}(v \oplus m)$. Then $v \leq_k \varphi^e_{\mathcal{B}}(v \oplus m)$ since $v$ is a fixed point. Similarly $m \leq_k \varphi^e_{\mathcal{B}}(v \oplus m)$, so $v \oplus m \leq_k \varphi^e_{\mathcal{B}}(v \oplus m)$. By Proposition 7.9.3 there is a consistent maximal fixed point $m'$ above $v \oplus m$, that is, $v \oplus m \leq_k m'$. Then $m \leq_k v \oplus m \leq_k m'$. Since $m$ is maximal, $m = m'$, and so $v \leq_k v \oplus m \leq_k m$.
2. Let $v$ be a fixed point that is below every consistent maximal fixed point; we show $v$ is intrinsic. Let $w$ be an arbitrary consistent fixed point of $\varphi^e_{\mathcal{B}}$; we show $v \oplus w$ is consistent. Since $w$ is a fixed point, $w \leq_k \varphi^e_{\mathcal{B}}(w)$, so by Proposition 7.9.3 $w \leq_k m$ for some consistent maximal fixed point $m$. $v \leq_k m$ by hypothesis. Then $v \oplus w \leq_k m \oplus m = m$, so also $-m \leq_k -(v \oplus w)$. Since $m$ is consistent, $m \leq_k -m$. Then $(v \oplus w) \leq_k m \leq_k -m \leq_k -(v \oplus w)$, so $v \oplus w$ is consistent.

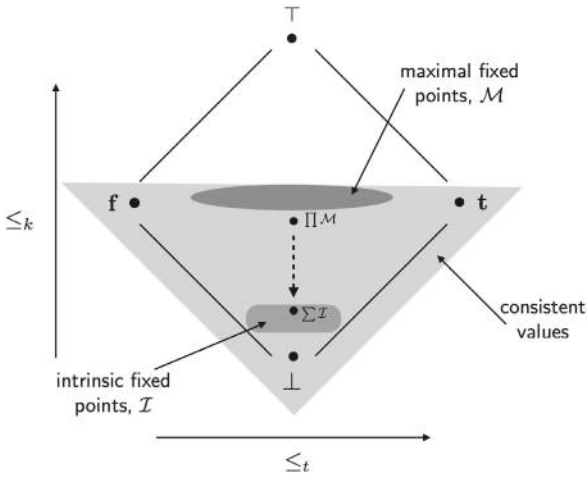



**Fig. 7.5** Maximal and intrinsic fixed points

□

In bilattices there is a largest as well as a smallest fixed point for truth revision operators. Confined to the consistent part there is no largest, though there are maximal ones. Confined further to the intrinsic fixed points there is a largest again, as was noted in Kripke (1975). The following includes some ideas from Fitting (1986) as well.

**Proposition 7.9.7** *There is a largest intrinsic fixed point of $\varphi^e_{\mathcal{B}}$. It is $\sum\mathcal{I}$ where $\mathcal{I}$ is the set of all intrinsic fixed points. An ordinal sequence constructed using $\varphi^e_{\mathcal{B}}$ as sketched in the proof of Proposition 7.5.2, starting at $\prod\mathcal{M}$, where $\mathcal{M}$ is the set of consistent maximal fixed points, proceeds downward in the $\leq_k$ ordering and converges to the largest intrinsic fixed point.*

***Proof*** Figure 7.5 shows the overall relationships between the various kinds of fixed points of $\varphi^e_{\mathcal{B}}$ that will be established here. Keep in mind that it is highly schematic. The bilattice with conflation that is shown is a space of valuations; only the extremal values are shown explicitly. All our work is confined to the consistent part of it, which is shown lightly shaded. Two subsets of the consistent valuations are shown shaded darker. $\mathcal{I}$ is the set of intrinsic fixed points, which includes the least fixed point. $\mathcal{M}$ is the set of consistent maximal fixed points.

First we show that $\sum\mathcal{I} \leq_k \prod\mathcal{M}$. Let $v$ be an arbitrary member of $\mathcal{I}$. Since $v$ is an intrinsic fixed point of $\varphi^e_{\mathcal{B}}$ then $v \leq_k \prod\mathcal{M}$ by Proposition 7.9.6. Since $v$ was arbitrary, $\sum\mathcal{I} \leq_k \prod\mathcal{M}$.

Next we show that if $v$ is a fixed point of $\varphi^e_{\mathcal{B}}$ and $\sum\mathcal{I} \leq_k v \leq_k \prod\mathcal{M}$ then $v$ is in fact the largest intrinsic fixed point and $v = \sum\mathcal{I}$. Very simply, $v$ is intrinsic by

Proposition 7.9.6, and $v$ is largest because, if $w$ is any intrinsic fixed point, $w \in \mathcal{I}$ and so $w \leq_k \sum \mathcal{I} \leq_k v$. Finally since $v$ is intrinsic, $v \in \mathcal{I}$ so $v \leq_k \sum \mathcal{I}$, and so $v = \sum \mathcal{I}$.

Finally we show a fixed point between $\sum \mathcal{I}$ and $\prod \mathcal{M}$ exists. This can be done in two different ways, and we sketch both.

For the first way, $\sum \mathcal{I}$ is a post-fixed point of $\varphi^e_{\mathcal{B}}$ by the following argument. For every $w \in \mathcal{I}$ we have $w \leq_k \sum \mathcal{I}$ so by monotonicity, $\varphi^e_{\mathcal{B}}(w) \leq_k \varphi^e_{\mathcal{B}}(\sum \mathcal{I})$, and so $w \leq_k \varphi^e_{\mathcal{B}}(\sum \mathcal{I})$ since $w$ is a fixed point. Since $w$ was arbitrary, $\sum \mathcal{I} \leq_k \varphi^e_{\mathcal{B}}(\sum \mathcal{I})$. Then by Proposition 7.5.2 part 2, there is a fixed point that is least above $\sum \mathcal{I}$ and, using Corollary 7.5.4, it is easy to show it is below $\prod \mathcal{M}$.

For the second way, $\prod \mathcal{M}$ is a pre-fixed point of $\varphi^e_{\mathcal{B}}$, with a proof dual to that for $\sum \mathcal{I}$ above. Using Proposition 7.5.2 part 3, there is a fixed point that is greatest below $\prod \mathcal{M}$ and, using Corollary 7.5.4 dualized, one shows it is above $\sum \mathcal{I}$.

The first of the two proofs just sketched involves an ordinal sequence that, ostensibly, moves upward from $\sum \mathcal{I}$, but in fact it is actually constant. The second argument uses an ordinal sequence that proceeds downward from $\prod \mathcal{M}$, approximating to $\sum \mathcal{I}$ from above. This is suggested by the dashed arrow in the figure. □

### *7.9.3 Alternating Fixed Points*

A very different class of fixed points was introduced in Yablo (1985), making use of *alternating* fixed points, Definition 7.5.5. These are closely related to the stable model semantics of logic programming, Gelfond and Lifschitz (1988), Fine (1989), and to well-founded semantics, Van Gelder (1989); Van Gelder et al. (1988, 1991). The treatment here is derived from Fitting (1993) on the logic programming side, and Fitting (1997) on the philosophical logic side.

A truth revision operator is monotone with respect to the $\leq_k$ ordering, but not in the $\leq_t$ ordering because of negation. It is possible, however, to localize this problem. Sentences of language $\mathcal{L}$ can have their negations pushed in to the atomic level using the various De Morgan laws. If this has been done we can think of occurrences of $\neg\mathsf{T}(x)$ as if they were occurrences of a new atom, a falsehood atom, that can behave independently of $\mathsf{T}(x)$.

**Terminology Convention 7.9.8** For this subsection only, a sentence must be in negation normal form, and so all occurrences of the negation symbol are at the atomic level.

The following introduces a mapping taking two valuations as input and producing what we call a *pseudo valuation* as output, where this independently maps positive and negative literals involving $\mathsf{T}$ to bilattice values.

**Definition 7.9.9** (Pseudo-Valuations) $\mathcal{A}(\neg\mathsf{T})$ is the set of negated atomic sentences in the language $\mathcal{L}$ of the form $\neg\mathsf{T}(0^{++\cdots+})$. The map $\triangle : \mathcal{B}^{\mathcal{A}(\mathsf{T})} \times \mathcal{B}^{\mathcal{A}(\mathsf{T})} \rightarrow \mathcal{B}^{\mathcal{A}(\mathsf{T}) \cup \mathcal{A}(\neg\mathsf{T})}$ is defined by the following conditions (where we write $\triangle$ in infix position):

$$(v_1 \triangle v_2)(\mathsf{T}(0^{++\cdots+})) = v_1(\mathsf{T}(0^{++\cdots+}))$$
$$(v_1 \triangle v_2)(\neg\mathsf{T}(0^{++\cdots+})) = \neg v_2(\mathsf{T}(0^{++\cdots+})).$$

$(v_1 \triangle v_2)$ is the *pseudo-valuation* of valuations $v_1$ and $v_2$.

We had various ways of extending valuations to all formulas, and it is the same for pseudo-valuations. The following is the counterpart of Definitions 7.6.2 and 7.6.4.

**Definition 7.9.10** (Pseudo-Valuation Extension) A *pseudo-valuation extension* is a mapping $\triangle^e : \mathcal{B}^{\mathcal{A}(\mathsf{T})} \times \mathcal{B}^{\mathcal{A}(\mathsf{T})} \to \mathcal{B}^{\mathcal{L}}$ (where $\mathcal{L}$ is here the set of sentences in negation normal form). The *monotonicity/anti-monotonicity property* holds for $\triangle^e$ if:

1. $k$ Monotonicity in Both Inputs: $v_1 \leq_k v_2$ and $w_1 \leq_k w_2$ implies $(v_1 \triangle^e w_1) \leq_k (v_2 \triangle^e w_2)$.
2. $t$ Monotonicity in First Input: $v_1 \leq_t v_2$ implies $(v_1 \triangle^e w) \leq_t (v_2 \triangle^e w)$.
3. $t$ Anti-Monotonicity in Second Input: $w_1 \leq_t w_2$ implies $(v \triangle^e w_1) \geq_t (v \triangle^e w_2)$.

Both strong Kleene and weak Kleene provide obvious pseudo-valuation extensions. For instance, Definition 7.8.7 for weak Kleene valuations, becomes the following, in bilattice $\mathcal{B}$.

1. For $L$ being an atomic sentence or its negation, involving $\mathsf{T}$, $(v_1 \triangle^w v_2)(L) = (v_1 \triangle v_2)(L)$.
2. For numerals $a$, $b$, and $c$:
   a. $(v_1 \triangle^w v_2)(A(a,b,c)) = v_1^w(A(a,b,c)) = v_2^w(A(a,b,c))$;
   b. $(v_1 \triangle^w v_2)(M(a,b,c)) = v_1^w(M(a,b,c)) = v_2^w(M(a,b,c))$;
   c. $(v_1 \triangle^w v_2)(a = b) = v_1^w(a = b) = v_2^w(a = b)$.
3. $(v_1 \triangle^w v_2)(X \wedge Y) = (v_1 \triangle^w v_2)(X) \wedge^w (v_1 \triangle^w v_2)(Y)$
4. $(v_1 \triangle^w v_2)(X \vee Y) = (v_1 \triangle^w v_2)(X) \vee^w (v_1 \triangle^w v_2)(Y)$
5. There is no separate negation case since formulas are in negation normal form.
6. $(v_1 \triangle^w v_2)(\forall x F(x)) = \bigwedge^w \{(v_1 \triangle^w v_2)(F(t)) \mid t \text{ is a numeral}\}$
7. $(v_1 \triangle^w v_2)(\exists x F(x)) = \bigvee^w \{(v_1 \triangle^w v_2)(F(t)) \mid t \text{ is a numeral}\}$

If language $\mathcal{L}$ is modal, we add the following, where $\mathcal{F} = \langle \mathcal{G}, \mathcal{R} \rangle$ is a modal frame.

8. $(v_1 \triangle^w v_2)(\Box X)(\Gamma) = \bigwedge^w \{(v_1 \triangle^w v_2)(X)(\Delta) \mid \Delta \in \mathcal{G} \text{ and } \Gamma \mathcal{R} \Delta\}$
9. $(v_1 \triangle^w v_2)(\Diamond X)(\Gamma) = \bigvee^w \{(v_1 \triangle^w v_2)(X)(\Delta) \mid \Delta \in \mathcal{G} \text{ and } \Gamma \mathcal{R} \Delta\}$

Similarly for strong Kleene. We omit details. Pseudo-valuation extensions of strong Kleene and weak Kleene both meet the monotonicity/anti-monotonicity conditions. This is straighhtforward to check, and we leave it to the reader. The status of pseudo-valuation extensions for supervaluations has not been investigated.

The following is the counterpart of Definition 7.6.3, for Truth Revision Operators.

**Definition 7.9.11** (Two Input Truth Revision Operator) Let $\triangle$ be a pseudo-valuation and $\triangle^e$ be an extension. Then $\psi^e_{\mathcal{B}}(v_1, v_2) : \mathcal{B}^{\mathcal{A}(\mathsf{T})} \times \mathcal{B}^{\mathcal{A}(\mathsf{T})} \to \mathcal{B}^{\mathcal{A}(\mathsf{T})}$ is the mapping such that for $\mathsf{T}$ valuations $v_1$, $v_2$ in $\mathcal{B}$,

$\psi^e_{\mathcal{B}}(v_1, v_2)(\mathsf{T}(\vDash X^\neg)) = (v_1 \triangle^e v_2)(X)$ for every negation normal form sentence $X$.

The following is an easily proved counterpart to Proposition 7.6.5.

**Proposition 7.9.12** *If the $\triangle^e$ monotonicity/anti-monotonicity property holds (Definition 7.9.10). Then the two input truth revision operator $\psi^e_{\mathcal{B}}$ has the following properties.*

1. *k Monotonicity in Both Inputs: if $v_1 \leq_k v_2$ and $w_1 \leq_k w_2$ then $\psi^e_{\mathcal{B}}(v_1, w_1) \leq_k \psi^e_{\mathcal{B}}(v_2, w_2)$.*
2. *t Monotonicity in First Input: if $v_1 \leq_t v_2$ then $\psi^e_{\mathcal{B}}(v_1, w) \leq_t \psi^e_{\mathcal{B}}(v_2, w)$.*
3. *t Anti-Monotonicity in Second Input: if $w_1 \leq_t w_2$ then $\psi^e_{\mathcal{B}}(v, w_1) \geq_t \psi^e_{\mathcal{B}}(v, w_2)$.*

The original motivation comes from logic programming. Make a guess at negative values, somehow choosing a valuation $v$ to tell us how occurrences of $\neg\mathsf{T}(t)$ behave. Relative to this guess, the least fixed point of $\lambda x.\psi^e_{\mathcal{B}}(x, v)$ tells a plausible way that positive $\mathsf{T}(t)$ atoms could then behave. If this turns out to be the guess we began with, then $v$ was a good guess. Here the $\leq_t$ ordering plays a fundamental role. This suggest the following analog of Truth Revision Operators, Definition 7.6.3.

**Definition 7.9.13** (Derived Operator) The *derived operator* of the two input truth revision operator $\psi^e_{\mathcal{B}}$ is the single input function $\widehat{\psi^e_{\mathcal{B}}}$ where $\widehat{\psi^e_{\mathcal{B}}}(v)$ is the smallest fixed point, in the $\leq_t$ ordering, of the function $\lambda x.\psi^e_{\mathcal{B}}(x, v)$.

**Proposition 7.9.14** *Assuming monotonicity/anti-monotonicity for $\triangle^e$, the function $\widehat{\psi^e_{\mathcal{B}}}$ is well-defined, is monotonic in $\leq_k$, and anti-monotonic in $\leq_t$.*

***Proof*** Assume $\psi^e_{\mathcal{B}}$ has the properties specified in Proposition 7.9.12.

$\widehat{\psi^e_{\mathcal{B}}}$ is well-defined. The smallest fixed point required in the definition of $\widehat{\psi^e_{\mathcal{B}}}$ exists because $\psi^e_{\mathcal{B}}$ is monotonic in its first input under $\leq_t$.

Monotonicity in the $\leq_k$ ordering. Assume $v_1 \leq_k v_2$. We define two ordinal sequences $F$ and $G$ by transfinite recursion. Both sequences consist of $\mathsf{T}$ valuations, members of $\mathcal{B}^{\mathcal{A}(\mathsf{T})}$. Also $\mathbf{f}$ is the $\mathsf{T}$ valuation that takes the value $\mathbf{f} \in \mathcal{B}$ on every input, $\alpha + 1$ is an arbitrary successor ordinal, and $\lambda$ is an arbitrary limit ordinal.

$$\begin{array}{rlrl} F(0) &= \mathbf{f} & G(0) &= \mathbf{f} \\ F(\alpha+1) &= \psi^e_{\mathcal{B}}(F(\alpha), v_1) & G(\alpha+1) &= \psi^e_{\mathcal{B}}(F(\alpha), v_2) \\ F(\lambda) &= \bigvee\{F(\alpha) \mid \alpha < \lambda\} & G(\lambda) &= \bigvee\{G(\alpha) \mid \alpha < \lambda\} \end{array}$$

From the proof of Proposition 7.5.2, both ordinal sequences are increasing, and both are constant from some point on, the first settling on the least fixed point of $\lambda x.\psi^e_{\mathcal{B}}(x, v_1)$ in the $\leq_t$ ordering, and hence to $\widehat{\psi^e_{\mathcal{B}}}(v_1)$, and the second to $\widehat{\psi^e_{\mathcal{B}}}(v_2)$. To prove $\widehat{\psi^e_{\mathcal{B}}}(v_1) \leq_k \widehat{\psi^e_{\mathcal{B}}}(v_2)$ we show that, for each ordinal $\alpha$, $F(\alpha) \leq_k G(\alpha)$. $F(\alpha) \leq G(\alpha)$ is trivially true for $\alpha = 0$. If it is true for all $\alpha < \lambda$, a limit ordinal, it is true for $\lambda$ using infinitary interlacing. This leaves the successor step. Suppose $F(\alpha) \leq_k G(\alpha)$. Then $F(\alpha+1) = \psi^e_{\mathcal{B}}(F(\alpha), v_1) \leq_k \psi^e_{\mathcal{B}}(G(\alpha), v_2) = G(\alpha+1)$ using $k$ monotonicity in both inputs.

Anti-monotonicity in the $\leq_t$ ordering. Assume $v_1 \leq_t v_2$; show $\widehat{\psi^e_{\mathcal{B}}}(v_2) \leq_t \widehat{\psi^e_{\mathcal{B}}}(v_1)$. $\psi^e_{\mathcal{B}}$ is anti-monotonic in its second argument under $\leq_t$, so $\psi^e_{\mathcal{B}}(\widehat{\psi^e_{\mathcal{B}}}(v_1), v_2) \leq_t \psi^e_{\mathcal{B}}(\widehat{\psi^e_{\mathcal{B}}}(v_1), v_1)$. Then $\psi^e_{\mathcal{B}}(\widehat{\psi^e_{\mathcal{B}}}(v_1), v_2) \leq_t \widehat{\psi^e_{\mathcal{B}}}(v_1)$ because $\widehat{\psi^e_{\mathcal{B}}}(v_1)$ is a fixed point of $\lambda x.\psi^e_{\mathcal{B}}(x, v_1)$. But then $\widehat{\psi^e_{\mathcal{B}}}(v_1)$ is a pre-fixed point of $\lambda x.\psi^e_{\mathcal{B}}(x, v_2)$ in the $\leq_t$ ordering, so there is a fixed point of $\lambda x.\psi^e_{\mathcal{B}}(x, v_2)$, call it $a$, below $\widehat{\psi^e_{\mathcal{B}}}(v_1)$. Since $\widehat{\psi^e_{\mathcal{B}}}(v_2)$ is the *least* fixed point of $\lambda x.\psi^e_{\mathcal{B}}(x, v_2)$, then $\widehat{\psi^e_{\mathcal{B}}}(v_2) \leq_t a \leq_t \widehat{\psi^e_{\mathcal{B}}}(v_1)$. □

**Definition 7.9.15** (Stable Fixed Points) Let $\psi^e_{\mathcal{B}}$ be a two input truth revision operator, and let $\widehat{\psi^e_{\mathcal{B}}}$ be the corresponding derived operator. We call fixed points of $\widehat{\psi^e_{\mathcal{B}}}$ *stable e* fixed points.

Monotonicity of derived operators in the $\leq_k$ ordering tells us that stable fixed points exist. Assuming reasonable conditions, stable fixed points are a subclass of the class of fixed points we have been looking at throughout.

**Definition 7.9.16** (Connected Condition) We say pseudo-valuation extension $e$ and valuation extension $e$ are *connected* if, for each $\mathsf{T}$ valuation $v$ and each sentence $X$ (in negation normal form, remember), $(v \triangle^e v)(X) = v^e(X)$.

Both strong Kleene and weak Kleene have connected valuation and pseudo-valuation extensions. This is easy, and we leave it to the reader.

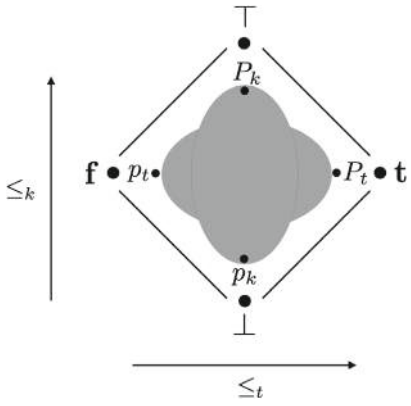
**Proposition 7.9.17** *Assume the monotonicity/anti-monotonicity conditions hold, and pseudo-valuation extension e and valuation extension e are connected. Then every fixed point of $\widehat{\psi^e_{\mathcal{B}}}$ is also a fixed point of $\varphi^e_{\mathcal{B}}$ (Definition 7.6.3). That is, every stable fixed point is a fixed point in the sense of Kripke.*

***Proof*** It follows easily from the definition of being connected that, for each $\mathsf{T}$ valuation $v$, $\varphi^e_{\mathcal{B}}(v) = \psi^e_{\mathcal{B}}(v, v)$. Also for each valuation $v$, $\widehat{\psi^e_{\mathcal{B}}}(v)$ is a fixed point of $\lambda x.\psi^e_{\mathcal{B}}(x, v)$ so $\psi^e_{\mathcal{B}}(\widehat{\psi^e_{\mathcal{B}}}(v), v) = \widehat{\psi^e_{\mathcal{B}}}(v)$. Now suppose $s$ is any fixed point of $\widehat{\psi^e_{\mathcal{B}}}$. Then $\varphi^e_{\mathcal{B}}(s) = \psi^e_{\mathcal{B}}(s, s) = \psi^e_{\mathcal{B}}(\widehat{\psi^e_{\mathcal{B}}}(s), s) = \widehat{\psi^e_{\mathcal{B}}}(s) = s$. □

Stable fixed points are thus among the family of Kripke style fixed points, but they are a *proper* subset. With strong Kleene, and using a Kripke style operator, fixed points can give the truth teller any value from the bilattice. This is not the case with stable fixed points. The definition uses the smallest fixed point in the truth ordering when evaluating the derived operator, so an explicit bias towards falsehood has been introduced, and in a stable strong Kleene fixed point the truth teller can only be **f**.

We know from Proposition 7.9.14 that the derived operator $\widehat{\psi^e_{\mathcal{B}}}$ is monotone under $\leq_k$ and anti-monotone under $\leq_t$. Because of monotonicity it has least and greatest fixed points under $\leq_k$, call them $p_k$ and $P_k$. Then, of course, all stable fixed points are between $p_k$ and $P_k$ in the information ordering, with the extremes included. Further, because of anti-monotonicity under $\leq_t$ the operator $\widehat{\psi^e_{\mathcal{B}}}$ has an extremal alternating fixpoint pair by Proposition 7.5.6, call the members of the pair $p_t$ and $P_t$. While these constitute an alternating fixpoint pair, they are not themselves fixed points. But

**Fig. 7.6** Stable fixed points




each fixed point can be thought of as an alternating fixpoint pair, with both members of the pair being the same, and so all stable fixed points are strictly between $p_t$ and $P_t$ in the truth ordering. The space of stable fixed points is shown schematically in Fig. 7.6. Further, the four extremal values are tightly connected; by Proposition 7.5.9: $P_t \otimes p_t = p_k$, $P_t \oplus p_t = P_k$, $P_k \wedge p_k = p_t$, $P_k \vee p_k = P_t$.

In Fitting (1997) the notion of intrinsic (Sect. 7.9.2) was relativized to stable fixed points. A stable fixed point $v$ is *stable intrinsic* if $v \oplus w$ is consistent for every consistent stable fixed point valuation $w$. There are stable intrinsic fixed points, but essentially no investigation of their properties has been made.

## 7.10 Conclusion

Several versions of theories of truth have been examined, in very general settings. Still, there could be more. For example, Kripke (1975, 706) made some interesting suggestions.

> The construction could be generalized so as to allow more notation in $\mathcal{L}$ than just first-order logic. For example, we could have a quantifier meaning "for uncountably many $x$," a "most" quantifier, a language with infinite conjunctions, etc. There is a fairly canonical way, in the Kleene style, to extend the semantics of such quantifiers and connectives so as to allow truth-value gaps, but we will not give details.

As far as I know, this has not been followed up in the original three valued setting. It certainly has not been using bilattice generalizations. It seems to be an open invitation still today. Kripke also mentioned including a modal operator. In the present treatment this is simply working with a bilattice of a special form, as in Example 7.4.9. One might carry this further and allow multiple knowledge modalities. Perhaps even communication between agents could be modeled. Then, of course, we could model the quirks of self referential truth that arise in the discussions of a

group of philosophers. (This is not an entirely frivolous suggestion.) We hope the present paper enables possibilities, and does not represent the end of the discussion.

It has been suggested that some detailed comments on the philosophical implications of the work presented here might be given, but I have chosen not to do this. I feel I have presented a rich set of tools. It is hoped that philosophical investigators will find some of the tools useful. But I am a theorist, and applications are not really my job.

## References

Arieli, O., & Avron, A. (1998). The value of four values. *Artificial Intelligence, 102*, 97–141.

Avron, A. (1996). The structure of interlaced bilattices. *Mathematical Structures in Computer Science*, *6*(3), 287–299. https://doi.org/10.1017/S0960129500001018

Belnap, N. D., Jr. (1977). A useful four-valued logic. In J. M. Dunn & G. Epstein (Eds.), *Modern uses of multiple-valued logic* (pp. 5–37). Netherlands: Springer.

Bochvar, D. A. (1938). Ob odnom trechznacnom iscislenii i ego primenenii k analizu paradoksov klassiceskogo rassirennogo funkcional'nogo iscislenija. *Matematiceskij Sbornik, 4*, 287–308. (English translation in Bochvar (1981).)

Bochvar, D. A. (1981). On a three-valued logical calculus and its application to the analyses of the paradoxes of the classical extended functional calculus. *History and Philosophy of Logic, 2*, 87–112. (Merrie Bergmann translator.)

Cobreros, P. (2013). Vagueness: Subvaluationism. *Philosophy Compass, 8*(5), 472–485.

Davey, B. (2013). The product representation theorem for interlaced pre-bilattices: Some historical remarks. *Algebra Universalis, 70*, 403–409.

Ferguson, T. (2015). Cut-down operations on bilattices. In: 2015 IEEE International Symposium On Multiple-valued Logic (pp. 24–29). https://doi.org/10.1109/ISMVL.2015.14

Fine, K. (1989). The justification of negation as failure. In J. E. Fenstad, I. T. Frolov, & R. Hilpinen (Eds.), *Logic, methodology and philosophy of science VIII* (pp. 263–301). Amsterdam: North-Holland.

Fitting, M. C. (1986). Notes on the mathematical aspects of Kripke's theory of truth. *Notre Dame Journal of Formal Logic, 27*, 75–88.

Fitting, M. C. (1989). Bilattices and the theory of truth. *Journal of Philosophical Logic, 18*, 225–256.

Fitting, M. C. (1990). Bilattices in logic programming. In G. Epstein (Ed.), *The twentieth international symposium on multiple-valued logic* (pp. 238–246).

Fitting, M. C. (1991). Bilattices and the semantics of logic programming. *Journal of Logic Programming, 11*, 91–116.

Fitting, M. C. (1993). The family of stable models. *Journal of Logic Programming, 17*, 197–225.

Fitting, M. C. (1994). Kleene's three-valued logics and their children. *Fundamenta Informaticae, 20*, 113–131.

Fitting, M. C. (1997). A theory of truth that prefers falsehood. *Journal of Philosophical Logic, 26*, 477–500.

Fitting, M. C. (2006). Bilattices are nice things. In T. Bolander, V. Hendricks, & S. A. Pederrsen (Eds.), *Self-Reference* (pp. 53–77). Center for the Study of Language and Information.

Gargov, G. (1999). Knowledge, uncertanity and ignorance in logic: Bilattices and beyond. *Journal of Applied Non-Classical Logics, 9*(2–3), 195–283.

Gelfond, M. & Lifschitz, V. (1988). The stable model semantics for logic programming. In R. Kowalski & K. Bowen (Eds.), *Proceedings of the fifth logic programming symposium* (pp. 1070–1080). Cambridge, MA: MIT Press.

Ginsberg, M. L. (1988). Multivalued logics: A uniform approach to reasoning in artificial intelligence. *Computational Intelligence, 4*, 265–316.
Kleene, S. C. (1950). *Introduction to metamathematics*. Princeton, NJ: D. Van Nostrand.
Kripke, S. (1975). Outline of a theory of truth. *The Journal of Philosophy, 72*(19), 690–716. (Reprinted in *New Essays on Truth and the Liar Paradox*, R. L. Martin, ed., Oxford (1983).)
Manna, Z. & Shamir, A. (1977). The optimal approach to recursive programs. *Communication of ACM, 20*, 824–831.
Martin, R. L. & Woodruff, P. W. (1975). On representing "true-in-L" in L. *Philosophia, 5*, 217–221. (Reprinted in *Recent Essays on Truth and the Liar paradox*, R. L. Martin editor, Oxford, 1984.)
Pynko, A. P. (2000). Regular bilattices. *Journal of Applied Non-Classical Logics, 10*, 93–111.
Smullyan, R. M. (1968). *First-order logic*. Berlin: Springer. (Revised Edition, Dover Press, New York, 1994.)
Szmuc, D. J. (2018). Track-down operations on bilattices. In R. Wille & M. Lukac (Eds.), *Proceedings of the 48th IEEE international symposium on multiple-valued logic* (pp. 74–79).
Teijeiro, P. (2020). Subvaluationism and classical recapture. *Logic Journal of the IGPL, 28*, 832–844. https://doi.org/10.1093/jigpal/jzy062
Van Fraassen, B. (1966). Singular terms, truth-value gaps, and free logic. *Journal of Philosophy, 63*, 481–485.
Van Gelder, A. (1989). The alternating fixpoint of logic programs with negation. In *Proceedings of 8th ACM symposium on principles of database systems* (pp. 1–10). Philadelphia: ACM.
Van Gelder, A., Ross, K. A., & Schlipf, J. S. (1988). Unfounded sets and well-founded semantics for general logic programs. In *Proceedings of Seventh Symposium on Principles of Database Systems* (pp. 221–230).
Van Gelder, A., Ross, K. A., & Schlipf, J. S. (1991). The well-founded semantics for general logic programs. *JACM, 38*, 620–650.
Visser, A. (1984). Four valued semantics and the liar. *Journal of Philosophical Logic, 13*, 181–212.
Yablo, S. (1985). Truth and reflection. *Journal of Philosophical Logic, 14*, 297–349.

# Chapter 8
# Logics for Rigidity

**James W. Garson**

**Abstract** Kripke is famous for holding that when the identity sign is flanked by proper names or natural kind terms, then the result is necessary if true. His conclusion is supported by the idea that proper names and natural kinds terms are rigid designators. To explore the cogency of Kripke's position, this paper takes on two interlocking projects in the formulation of the semantics for quantified modal logic (QML). Project 1. Define 'rigidity' in a way that is faithful to Kripke's intentions, and for which evidence for the rigidity of proper names and natural kind terms is forthcoming. Project 2. Given a candidate definition of rigidity produced in Project 1, find a well-motivated semantics for QML that validates the claim that 't is identical to s' is necessary if true when terms t and s are rigid in that sense. It is surprisingly difficult to succeed at both projects. Three different definitions for 'rigidity' are explored here along with a variety of corresponding QMLs. The few QMLs that survive Project 2 are very different from ones Kripke developed in his seminal paper "Semantic Considerations in Modal Logic". Furthermore, there are serious costs whichever solution we attempt.

**Keywords** Free logic · Natural kind terms · Necessary identity · Quantified modal logic · Proper names · Rigidity

J. W. Garson (✉)
Department of Philosophy, University of Houston, Houston 7204 3004, USA
e-mail: jgarson@uh.edu

© The Author(s), under exclusive license to Springer Nature Switzerland AG 2024

Y. Weiss and R. Birman (eds.), *Saul Kripke on Modal Logic*, Outstanding Contributions to Logic 30, https://doi.org/10.1007/978-3-031-57635-5_8

## 8.1 Introduction

Kripke (1972/1980) is famous for holding that when t and s are proper names or natural kind terms and ≈ is the sign for identity, then t≈s is necessary if true,[1] despite the fact that, for the most part, such sentences are not a priori, but known instead by empirical evidence. The notion of a rigid designator is central for establishing this conclusion.[2] Two crucial premises are needed for the argument.

(1) Proper names (or natural kind terms) t and s are rigid designators.
(2) Whenever t and s are rigid, t≈s is necessary if true.

From a formal point of view, (2) amounts to the claim that (□ ≈) is valid in a predicate modal logic with identity, presuming that t and s are rigid.

(□ ≈) t≈s→ □t≈s

(In this paper we use 't' and 's' as abbreviations for proper names or natural kind terms. The logical constants are → (if..then), ~ (not), ∃ (the existential quantifier), ≈ (identity) and □ (necessity). The other logical constants: &, ∨, ↔, ∀, and ◇ are defined in the usual way.) At a first pass, the validity of (□ ≈), when t and s are rigid, seems an uncontroversial feature of modal logic. The rigidity of t guarantees (roughly) that t refers to the same object in every possible world, and similarly for s. But if the antecedent t≈s of (□ ≈) is true at any world, then t and s refer to the same thing in that world. By the rigidity of t and s, it follows immediately that t and s must refer to the same thing in every possible world, guaranteeing the truth of □s≈t.

However, the reasoning just given depends on details about how rigidity is defined, and how □ is interpreted. I used the parenthetical 'roughly' above because there are at least three different ways to spell out what Kripke intended by 'rigidity' in *Naming and Necessity* (Kripke, 1972/1980). Although the difference between them may appear to be of little consequence, it is the purpose of this paper to reveal devils in the details. The crucial connection Kripke needs between rigidity of t and s and

[1] This paper adopts the convention that identities involving natural kind terms s and t have the form: t≈s. Soames (2002, Chap. 9, pp. 246ff.) holds that this is a mistake, and that those terms are better symbolized with predicate letters. He also argues (on grounds different from what is presented here) that it is far from clear how to define rigidity for these expressions, and that the argument from rigidity to the necessity of true identities is not secure. This paper shows that even if Soames' issues about natural kind terms can be resolved, a more basic puzzle about rigidity must be faced—one that arises for proper names as well. For those who share Soames' views about natural kind terms, this paper may be taken as new evidence by *reductio* that he was right to worry, as our convention that natural kind terms are singular terms leads to a brand new collection of problems.

[2] Soames (2002, Chap. 10) argues that in the case of natural kind terms, Kripke's doctrines about necessary identities are better supported with evidence that does not involve claims of rigidity. Nonetheless, it is still of interest to explore the cogency of the widely held view that there is a straightforward entailment to be had between rigidity and those doctrines.

the validity of ($\Box \approx$) is not so easily secured, and when it is secured, there are costs to be faced.

To explore this issue, two interlocking projects in the formulation of the semantics for quantified modal logic (QML) will be undertaken.

Project 1. Define 'rigidity' in QML in a way that is faithful to Kripke's intentions, and for which evidence for the rigidity of proper names and natural kind terms is forthcoming.

Project 2. Given a candidate definition of rigidity produced in Project 1, find a well-motivated semantics for QML that validates ($\Box \approx$) when t and s are rigid in that sense.

It is surprisingly difficult to succeed at both projects. The best choices we have to solve Project 1 cause unexpected complications in Project 2. The QMLs that emerge are very different from the system that Kripke has developed in his seminal work "Semantic Considerations in Modal Logic" (Kripke, 1963). Kripke's system is conservative—it uses standard principles for the quantifiers and a two-valued treatment of truth. Instead, the QMLs that show any hope of satisfying Project 2 use non-standard rules for quantifiers, and/or truth-value gaps. In the latter case, new truth conditions for $\Box$ and a new definition of validity are needed to manage sentences that lack truth-values. Although it is possible to obtain a truth-value gap semantics for QML where rigidity of t and s ensure the validity of ($\Box \approx$), the cost is a counterintuitive reinterpretation of $\Box$ that appears incompatible with the position that Kripke is arguing for in *Naming and Necessity*.

Let us begin by describing this paper's narrative structure. The main character in our story is a person who is skeptical (as some reasonable philosophers are) about Kripke's conclusion that ($\Box \approx$): $t \approx s \rightarrow \Box t \approx s$ is valid. Call this person the Necessary Identities Skeptic, or the Skeptic for short. The Skeptic believes (for example) that there is a possible world where 'Hesperus' and 'Phosphorus' refer to different celestial bodies, or one where 'water' refers to XYZ, rather than $H_20$. So to give the Skeptic a fair hearing, we begin our story by allowing the possibility that terms t and s may refer to different things across the possible worlds. However, in the next episode, the Skeptic has heard a definition of 'rigidity' that satisfies Project 1—so good reasons were given to think that t and s are rigid in this sense. That means the Skeptic must accept a corresponding formal constraint on the intensions of these terms, which we express in the appropriate QML. The climax of the story is the evaluation of the validity status of ($\Box \approx$) in that QML. The denouement will prompt a reconsideration of the results obtained in Project 1, and this will begin another round.

## 8.2 Three Definitions of Rigidity

There are at least three different ways to define rigidity of a term t. The most straightforward is Simple Rigidity (or s-rigidity for short). The idea is that t is rigid iff t refers to the same object in every possible world. To give a more formal account, assume

that a model M includes a set W, of possible worlds, a domain O of objects, and an assignment function v that assigns to any term t and world w an object v(w, t) in O. Then simple rigidity may be defined as follows.

(Simple Rigidity) t is s-rigid iff there is a unique object o such that t refers to o in every possible world.

Simple Rigidity makes Project 2 easy, for when t and s are simply rigid, the validity of (□ ≈) can be shown following the informal explanation given in the previous section. To prove that result rigorously, we make explicit the truth conditions for ≈ and □ as follows.

(≈) v(w, s≈t) = T iff v(w, s) = v(w, t).
(□) v(w, □A) = T iff for all u in W, v(u, A) = T.

Here, the simplest account for □ was chosen, where the truth of □A amounts to truth in all possible worlds. Weaker truth conditions for □ using an accessibility relation could be substituted for (□), but that would not affect the results of this paper. With these truth conditions in place, a formal proof of the validity of (□ ≈): t≈s→ □t≈s follows as a simple exercise.

However, it does not seem that Kripke was comfortable with simple rigidity, for while this appears to be the literal reading of the text when 'rigidity' is first defined (as he says "quasi-technically") in *Naming and Necessity* (Kripke, 1972/1980, p. 48), he appears to immediately correct himself by suggesting that t is rigid iff t refers to the same object o in every possible world *where o exists*. Call this reading Existential Rigidity (or e-rigidity). What motive would Kripke have for choosing e-rigidity over s-rigidity?

Here is one possible answer (but an answer we will reconsider in Sect. 8.5 below). Suppose our language contains terms h and p for 'Hesperus' and 'Phosphorus'. Consider a model with a world w1 where Venus does not exist.[3] Simple rigidity requires that once h refers to Venus in the real world, it must refer to Venus in every possible world, including world w1. So s-rigidity entails that terms do not have existential import—that is, that it is possible for terms h and p to refer to objects that do not exist in w1. Kripke might have worried that the Skeptic disagrees, and holds the fairly widespread opinion that terms do have existential import. But if so, then convincing the Skeptic that proper names and natural kind terms are simply rigid is

[3] For the purposes of this paper, I have assumed that it makes sense to say that Venus does not exist in some possible worlds and that the domain Ow of objects existing in a world w may vary from one world to another. Some readers will challenge this. For example, advocates for the simplest QML (Linsky and Zalta, 1994; Williamson, 1998) adopt an ontology that includes a fixed domain of possibilia or other possibly non-concrete entities referred to by terms, so that 'Venus exists' turns out to be true in every world. Nevertheless, those who take this view make a distinction between possibilia that are realized, actual, or concrete and those that are not part of the spatio-temporal layout of a given possible world. Readers of this persuasion should understand that by 'Venus exists' I do not mean that Venus is among the possibilia, but rather that Venus is concrete. Concerns that adopting the simplest QML might lead to different conclusions from those initially drawn in this paper will be addressed in Sect. 8.6 below.

a hopeless project. If there are any worlds at all where Venus does not exist, then those are worlds where the referent of h cannot be Venus. The result is that proper names cannot be s-rigid unless their referents exist necessarily.

Another of Kripke's worries might be that if the domain for ∃ in w1 contains only what exists in w1, and h refers outside that domain, then ∃xx≈h, which asserts that what h refers to exists, will fail in w1. So ∃xx≈h (a theorem of predicate logic) is invalid, and the rules of quantification will need revision. That this is one of Kripke's concerns is supported by the pains he takes to preserve classical quantifier rules in "Semantical Considerations in Modal Logic" (Kripke, 1963).

So let us explore the reasonable assumption that by rigidity Kripke intended e-rigidity, and that places in the text where s-rigidity is literally expressed are to be understood as abbreviations for the more sophisticated reading. Right away there is trouble. It turns out that e-rigidity is totally useless for Project 2. The rider: 'where o exists' opens the door for counterexamples to (□ ≈).[4] Here are details.

To reflect e-rigidity in a formal semantics, assume that a model M also includes domains Ow (that are subsets of O) for each of the possible worlds w in W. Ow is understood to contain the things that exist in world w. Then Existential Rigidity (or e-rigidity) can be defined as follows.

(Existential Rigidity) t is e-rigid iff there is a unique object o such that t refers to o in every possible world w where o ∈ Ow.

A model M* that is a counterexample to (□ ≈), where h and p are e-rigid is just around the corner. Let its set of possible worlds W* contain r (for the real world), w1 and w2. Let O be {Venus, Vh, Vp}, where Vh and Vp are two different celestial objects neither identical to Venus. The domain Or for the real world contains Venus only and the domains for Ow1 and Ow2 contain both Vh and Vp, but not Venus. Let the valuation function v* for M* assign Venus to h and p in the real world r, so that v*(r, h≈p) = T. In w1, v* assigns to h and p the objects Vh and Vp, and in w2 these assignments are reversed: v*(w2, h) = Vp and v*(w2, p) = Vh. Since Venus only exists in r, h refers to Venus in all worlds where Venus exists. Furthermore, the varying referents of h across worlds w1 and w2 guarantees that h does not refer to Vh in every world where Vh exists, nor does it do so for Vp. So Venus is a unique object that h refers to in every world where Venus exists. That means h is e-rigid, and a similar argument establishes that p is e-rigid as well. Since there is a difference in the referents of h and p in w1, v*(w1, h≈p) = F. That means □h≈p must be rejected at r because there is a world where h≈p is false. In summary, v*(r, h≈p) = T and v*(r, □h≈p) = F, and so M* violates an instance of (□ ≈): t≈s→ □t≈s.

The problem is that e-rigidity does nothing to stop a model from rating h≈p false at worlds where the referent (Venus) of h and p in the real world does not exist. Those worlds are exempted from consideration using e-rigidity, and so nothing in

[4] This result is pointed out by Gomez-Torrente (2006, pp. 239–240). However, the counterexample they describe depends on the inclusion of a term that fails to refer to something that exists. In the countermodel M* to be presented below, all terms refer to things in the domain of quantification for each world. This is important in our case, since it was presumed that terms have existential import in the course of solving Problem 1.

the definition of e-rigidity stops us from setting the referents of h and p to different things in those worlds. Whatever Kripke's motivations for formulating e-rigidity might have been, it simply does not do what is required, namely to tie rigidity to the validity of ($\Box \approx$).

One might object that the account so far given for e-rigidity is too weak to capture what Kripke intended. The idea that a term like h is rigid in M*, when it refers to different things, is counterintuitive. A simple rider (italicized below) would appear to be needed to correct the definition of existential rigidity.

(Corrected Existential Rigidity) t is ce-rigid iff there is a unique object o such that t refers to o in every possible world where o $\in$ Ow, *and to nothing else*.[5]

The corrected definition better matches our intuitions about what rigidity means, and it solves the problem posed by the counterexample M*. Since h and p refer to things other than Venus in w1, neither qualifies as ce-rigid.

However, new complications arise. Let us presume for the moment that valuations are fully defined—they always assign a value to each term in each possible world. Then the rider: 'and to nothing else' ensures that ce-rigidity collapses to s-rigidity. Given that each term t must refer to something in each possible world, and given that t refers to nothing other than what t refers to in the real world, then t must refer to that same thing in every world. So t is s-rigid. But, Kripke clearly thought there was a difference between rigidity and s-rigidity in *Naming and Necessity*, so the corrected definition would appear to be incompatible with his intentions. Furthermore, if ce-rigidity amounts to s-rigidity, then the problem of dealing with the Skeptic who believes TEI—that terms only refer to existing objects—must be dealt with all over again. So far, ce-rigidity gets us nothing new.

The way out is to explore a semantics that allows non-referring terms. If h and p fail to refer in worlds where Venus does not exist, then TEI may be preserved. Furthermore, the distinction between s-rigidity and ce-rigidity is maintained because the former requires terms to be defined in each world, while the latter does not. Another advantage of reference-gap semantics is that the corrected version of e-rigidity can be recast in a more economical form, where the rigidity of t means that t refers to the same object in every world where t has any referent at all. Call this referential rigidity.

(Referential Rigidity) t is r-rigid iff there is a unique object o such that t refers to o in every possible world where t is defined.

Although it does not appear explicitly in the text of *Naming and Necessity*, r-rigidity appears to be an attractive candidate. If non-referring terms are allowed in the semantics, it is a straightforward matter to show that ce-rigidity and r-rigidity amount to the same thing.[6] Therefore, let us explore the idea that Kripke or his

[5] Soames (2002, p. 4) employs this rider in his definition of rigidity without comment. However, I have not found any place in *Naming and Necessity* where Kripke explicitly states the rider.

[6] Here are more details on why r-rigidity and ce-rigidity are equivalent. Suppose t is r-rigid. Then it follows that whenever t refers, t refers to the same object o in each world w. Therefore it refers to

defender, might choose r-rigidity as the definition of rigidity, and then prove the validity of ($\Box \approx$) in a QML whose semantics allows non-referring terms.[7] A lot will depend on exactly how such a QML is defined. Several methods for handling gaps come to mind, including three-valued logics and systems using super-valuations (Prior, 1957; Hintikka, 1961; van Fraassen, 1969). Canvassing all these options is beyond the scope of this paper. However, Gabbay (1976, pp. 75ff) has produced an elegant and accessible version of a semantics for QML that is well-equipped to the task at hand. This framework, called statability semantics, will help illustrate the options available, the issues to be resolved, and the related costs and benefits.

## 8.3 Statability Semantics

It seems the best hope for resolving the problem posed in this paper is to use models that allow non-referring terms. That means that we must define models where the valuation function v is *partial*, so that it may fail to assign a referent to some terms in some worlds. It is natural to hold that sentences containing such terms lack truth-values, so that valuations need not assign truth-values to every sentence in each world.[8] When a term t or sentence A fails to refer at world w, we write: 'v(w, t) is undefined' or 'v(w, A) is undefined'.

The use of partial valuations raises new questions concerning how to construct the truth conditions for sentences that allow gaps. Assume an *s-model* M contains a set of possible worlds W, a domain O of possible objects, domains Ow for each world w ∈ W that are subsets of O, and a partial valuation function v such that when v is defined for t at w, v(w, t) ∈ O,[9] and when v is defined for A at w, v(w, A) is a truth-value: T or F. Let A be *statable* in w on s-model M iff M's valuation function v assigns values in w to all terms appearing in A. In case TEI is assumed, M will obey the condition that each term t refers in w only when its value is a member of Ow. In that case, terms occurring in any sentence statable in w must refer to things that exist in w.

---

the same object o in all worlds w where o ∈ Ow, and the first conjunct of the definition of ce-rigidity is satisfied. The rider of that definition is also satisfied since r-rigidity ensures that t cannot refer to anything other than o. Now suppose t is ce-rigid. The rider guarantees that when t refers, it can only refer to the same object across the possible worlds. So t is r-rigid.

[7] The notion of r-rigidity is slightly weaker than persistence (Salmon, 1981, pp. 33–34). The reason is that a term can be r-rigid even if it happens to pick out the same object o in some worlds where o does not exist. But persistence makes the additional demand that a term must be undefined under those circumstances. We have chosen r-rigidity over persistence in this paper to simplify the presentation. The difference does not matter in the end because the model M** used in the next section to show the invalidity of ($\Box \approx$) has terms that are persistent as well as r-rigid.

[8] Note that actually in Gabbay's semantics, all terms refer. However, it is natural to relax that restriction (as we do here).

[9] For generality, Gabbay allows the possibility that v(w, t) ∉ Ow, and so terms need not have existential import. Since we want to explore the more general case where TEI need not hold, we follow his lead.

The truth value v(w, A) for A at w is defined iff A is statable in w on M. Furthermore, where v is defined in w for a formulas A and B, v meets the standard truth conditions for $\approx$, ~, $\rightarrow$, and $\exists$, which follow.

($\approx$) v(w, t$\approx$s) = T iff v(w, t) = v(w, s).
(~) v(w, ~A) = T iff v(w, A) = F.
($\rightarrow$) v(w, A$\rightarrow$B) = T iff v(w, A) = F or v(w, B) = T.
($\exists$) v(w, $\exists$xA) = T iff for some o $\in$ Ow, $v^o_x$(w, A) = T,
where $v^o_x$ is the valuation identical to v save that $v^o_x$(x) = o.

Following the same pattern, we might adopt the standard truth condition for $\Box$, assuming again that v does not define a value for $\Box$A unless $\Box$A is statable.

($\Box$) v(w, $\Box$A) = T iff v(u, A) = T, for every u in W.

Unfortunately, this choice has untoward consequences. There is a counterexample to ($\Box \approx$) for r-rigidity and it even satisfies TEI. Consider a model M** where h and p refer to Venus in the real world r. Presume as well that there is only one other world w1, where Venus does not exist, and h and p are not defined there. Then h and p are both r-rigid. But h$\approx$p is not statable in w1, since neither h nor p refers there. So v(w1, h$\approx$p) is not defined and that means v(w1, h$\approx$p) is not T. Since h and p are defined at r, $\Box$h$\approx$p is statable in r, and so we can apply ($\Box$) to determine whether $\Box$h$\approx$p is T or F in r. But for v(r, $\Box$h$\approx$p) to be T, it is required that v(u, h$\approx$p) = T for every world u. However, we had that v(w1, h$\approx$p) was not T. This means that that $\Box$h$\approx$p must be F at r. The counterexample has returned. The upshot is that to validate ($\Box \approx$), we need a different way to give the truth conditions for $\Box$.

Gabbay (1976, p. 75) explores what is, as far as I can tell, the only other reasonable option for what the truth condition for $\Box$ should be when there are truth-value gaps. The answer, which is used to define what he calls s-validity, takes its inspiration from the solution to a different puzzle in a semantics with truth-value gaps. The problem is is to find a way to validate theorems such as t$\approx$t and Ft$\rightarrow$Ft. The ordinary definition of validity of a sentence A requires that A be T at every world in every model. But v(w, Ft$\rightarrow$Ft) is not T when Ft$\rightarrow$Ft is not statable at w. The solution to this problem is to define validity so that a sentence qualifies as s-valid provided that it is true in every world *where it is statable*, on every s-model.

This revision to the concept of validity puts constraints on how we may interpret $\Box$. The rule of Necessitation, (which says that when A is a theorem so is $\Box$A) is a standard feature of normal modal logics. To guarantee that it preserves validity, it is a requirement that the s-validity of A entails the s-validity of $\Box$A. In the present context, that means that the necessity of A must not demand that A is true in all possible worlds but only in worlds where A is statable. This suggests the following revised truth condition ($\Box$s) for $\Box$.[10]

[10] Hintikka (1961, p. 125) proposes the analog of ($\Box$s) in his truth condition (C.N*) for $\Box$. He uses a truth-value semantics with the substitution interpretation for the quantifiers, and treats the possible words as state-descriptions in languages with varying collections of free variables. The free variables presumably pick out things that exist in a given world. So the analog of statability of

(□s) v(w, □A) = T iff v(u, A) = T for every u in W *where v(u, A) is statable*.

Call the semantics that uses (□s) and the other features of stability semantics that have been described so far: *s-semantics*.[11] A promising feature of s-semantics is that Problem 2 is solved: as (□ ≈): t≈s→ □t≈s comes out s-valid when t and s are r-rigid.[12] It appears that s-semantics provides everything Kripke could hope for, as it ties rigidity to the validity of (□ ≈) in the desired way.[13]

But despite these attractions of s-semantics, there are serious problems to face. It is interesting to note that Kripke mentions models with partial valuations in "Semantical Considerations" (Kripke, 1963, p. 86), but chose not to use them. He cited the work of Prior (1957) and Hintikka (1961) mentioning a cost for each—namely that Prior has to revise the modal propositional logic, and Hintikka has to restrict the rule of substitution.

There are costs for stability semantics as well. First, the laws of modal propositional logic have to be adjusted. The axiom of (Distribution) □(A→B)→(□A→ □B) must be restricted so that all terms and free variables in A must also appear (free) in B.[14] A second and deeper concern is to provide independent and convincing reasons for choosing (□s) over the more standard interpretation (□) of □. If (□) is

---

A in a world (or model set) w is that all free variables in A appear in formulas of w, and this is the idea he uses in (C.N*).

[11] Officially s-validity is defined as follows. An s-model M = <W, D, v> contains a set of possible worlds W, a domain structure D consisting of a set O of possible objects, and domains Ow for each world w ∈ W that are subsets of O, and a partial valuation function v such that when v is defined for t at w, v(w, t) ∈ O, and v(w, A) is defined exactly when A is statable at w, and when v is defined for A at w, v(w, A) is a truth-value: T or F, fixed by the truth conditions (≈), (~), (→), (∃), and (□s) above. Sentence A is s-valid iff v(w, A) = T whenever A is statable, in every world w for every s-model M=<W, D, v>. (For simplicity, predicate letters other than ≈ are omitted from the language.).

[12] Here is the proof. Consider any model where s and t are r-rigid. We show that t≈s→ □t≈s is true when statable in every world of every s-model. Suppose t≈s→ □t≈s is statable at w, for an arbitrary choice of world w in W. Hence v(w, t) and v(w, s) are both defined. Assume v(w, t≈s) = T. We will show that v(w, □t≈s) is both statable and T at w on this model. Since t≈s is statable at w, (≈) yields v(w, t) = v(w, s). To show that v(w, □t≈s) = T it suffices to show that v(u, t≈s) = T in all worlds u where t≈s is statable. So assume t≈s is statable in an arbitrarily selected world u*. That means that v(u*, t) and v(u*, s) are both defined. The r-rigidity of t ensures that v(u, t) = v(w, t) in all worlds u where t is defined, and similarly for s. But t and s are both defined at u*, so v(u*, t) = v(w, t) = v(w, s) = v(u*, s). But that means, by (≈), that v(u*, t≈s) = T. The argument was general with respect to the choice of u*, so v(u, t≈s) = T in all worlds u where t≈s is statable. Since t and s refer in w, □t≈s is statable in w, and it follows by (□s) that v(w, □t≈s) = T. It follows that t≈s→ □t≈s is statable in w, and by (→), T at w. Since our choice of model was arbitrary, we have that t≈s→ □t≈s is true when statable in every world of every s-model.

[13] One might wonder whether the s-validity of (□ ≈) might also be established for e-rigidity, as the use of (□s) in place of (□) might defeat the counterexample M*. It does not. The reason is that the model M* that qualified as a counterexample to (□ ≈) is an s-model where all sentences are statable. In that context, the effects of (□) and (□s) are the same.

[14] The reason for the restriction on (Distribution) can be explained using the following informal counterexample in a case where the restriction is violated. It provides a further illustration of the counterintuitive nature of the truth condition (□s) for □. Consider □(Cg→p)→(□Cg→ □p), where g abbreviates 'God', C abbreviates 'created Earth', and p abbreviates 'the Earth is perfect'. Here the restriction is violated since g occurs in Cg but not in p. Now consider a model with the following features. (1) It contains world w* where God exists. (2) In all worlds w of the model, g refers in

to be preferred, (□ ≈) fails and the Skeptic wins. It is hard to see what those reasons could be, in light of the fact that (□s) is at odds with our intuitions in ways that (□) is not.

To illustrate, consider a Skeptic who disagrees with Kripke, and believes that w (for: 'water' in English) means "functional-water", that is, whatever fills lakes and streams, and falls from the sky as rain, etc. That amounts to saying that w picks out $H_20$ in all worlds where water exists, but for worlds like Twin Earth (where there is no $H_20$) it refers to XYZ, or whatever (if anything) plays the functional role of $H_20$ on Earth. The Skeptic also believes (perhaps because of faith in TEI) that h (for: '$H_20$') always refers (when it does) to $H_20$, but only refers in worlds where $H_20$ exists. So h remains undefined on Twin Earth. The Skeptic's view flatly disagrees with Kripke's, for while Kripke allows 'water' to refer only to $H_20$ whenever it refers, the Skeptic thinks it refers to other (existing) things in worlds where $H_20$ does not exist. So for the Skeptic, 'water' is a not an r-rigid designator and the claim □w≈h[15] (for: water is necessarily $H_20$) should, by any reasonable intuitions, come out false.

The problem is that on the scenario described, (□s) rules that □w≈h is true! All it takes to verify □w≈h on (□s) is for 'water' to pick out $H_20$ in worlds where w≈h is statable. But those worlds are worlds where h (for $H_20$) refers—worlds where water exists. So all □w≈h says, according to (□s), is that 'water' picks out $H_20$ in worlds where $H_20$ exists, and this is something with which the Skeptic agrees. The truth condition (□s) systematically ignores all worlds where the Skeptic's view about 'water' and Kripke's differ, namely worlds where $H_20$ does not exist. Therefore, the validity of w≈h→ □w≈h is no threat to the Skeptic. He or she may cheerfully grant Kripke the validity of (□ ≈) on an interpretation (□s) of □ that is so bizarrely anemic that □w≈h is true when 'water' means functional water. Establishing that true identities are necessary in this weakened sense has nothing to do with what Kripke intended to argue for in *Naming and Necessity*.

The upshot is that the stability strategy is in deep trouble. As far as I can see, there are only two viable candidates for the truth condition for □. The standard one (□) simply does not validate (□ ≈), and while the alternative (□s) does, it requires modification of the principles of the underlying propositional modal logic, and it interprets □ in a counterintuitive way that trivializes Kripke's conclusion that true identities are necessary.

---

w exactly when God exists in w. (3) God creates the Earth in any world where he exists. (4) Every thing God creates is perfect. (5) There is a world w1 where God does not exist and the Earth is not perfect. By (2) and (3), Cg is T in exactly those worlds where Cg is statable. By (1), □Cg is statable and true in w*. By (1) and (4) □(Cg→p) is statable and T in w*. Since p does not contain g, □p is statable in w*. So □p is F there, because there is a possible world w1 with an imperfect Earth. Since g refers in w*, □(Cg→p)→(□Cg→ □p) is statable in w*. By the truth condition (→) for →, v(w*, □(Cg→p)→(□Cg→ □p)) = F. The counterintuitive nature of (□s) is underscored in this example by the fact that □Cg counts as true in w* despite the fact that there is a possible world w1 in this model where God does not create the Earth.

[15] According to the conventions of this paper, natural kind terms 'water' and '$H_2O$' are treated here as terms w and h rather than predicate letters. (See note 1.) If that bothers you, the same point can be made where proper names 'Hesperus' and 'Venus' replace 'water' and '$H_2O$' respectively.

## 8.4 Weakening the Concept of Necessity

There is another tactic worth considering for showing how rigidity could support Kripke's thesis concerning necessary identities. The idea might work for either referential or existential rigidity. The referential (existential) rigidity of terms t and s only guarantees that they refer to the same object in some of the possible worlds—the ones where they have a referent (where the referent exists). This fails to support ($\Box \approx$) because the truth of its consequent $\Box t \approx s$ requires that t and s refer to the same thing in all possible worlds. However, we might solve the problem by weakening this consequent so that there is a better match with the worlds that should be considered. Suppose we introduce the symbol E into the object language with the understanding that v(w, Et)=T iff v is defined for t at w (in the case of existential rigidity, v(w, Et)=T iff v(w, t) ∈ Ow). Now consider ($\Box \approx$E).

($\Box \approx$E) $\quad t \approx s \rightarrow \Box((Et \& Es) \rightarrow t \approx s)$

It is not difficult to show that ($\Box \approx$E) is indeed valid given the referential (existential) rigidity of t and s. So if Kripke's doctrine on necessary identities is correctly expressed by ($\Box \approx$E), then all is well. There is some textual evidence that this is what Kripke had in mind, for he tentatively endorses a version of ($\Box \approx$E) in *Naming and Necessity"* (Kripke, 1972/1980, p. 110).[16]

Unfortunately, this will not do. In the previous section we showed that weakening the truth conditions for $\Box$ to ($\Box$s) results in a version of ($\Box \approx$) that fails to express Kripke's intentions. Exactly the same problem emerges for ($\Box \approx$E). Consider again the Skeptic who believes that 'water' means functional water, so that it refers to XYZ in Twin Earth worlds where $H_20$ does not exist. Then ((Ew&Eh)→w≈h) (with w for 'water' and h for '$H_20$') is true in Twin Earth since Eh fails. That guarantees that w≈h→ $\Box$((Ew&Eh)→w≈h) is true under circumstances where w is a non-rigid designator! So the Skeptic can be perfectly happy with ($\Box \approx$E) since it in no way challenges his belief that 'water' means whatever fills lakes and streams, and falls from the sky as rain on Twin Earth. The moral of the story is exactly the same as for stability semantics. When necessity is weakened, Kripke's doctrines on necessary identities fail to be expressed.

## 8.5 Simple Rigidity Once More

Serious problems have been encountered for all three of the definitions of rigidity we have considered. In desperation, we might give simple rigidity (s-rigidity) a second look. Could it be that Kripke's mention of e-rigidity was just misguided,

[16] Actually, Kripke suggests replacing $\Box$h≈p with $\Box$(Eh→ h≈p) (with h for 'Hesperus' and p for 'Phosphorus'). Note that adding Eh in the antecedent without Ep as well is not sufficient to guarantee validity.

and s-rigidity is in fact the best choice for establishing that identities whose flanking terms are rigid are necessary if true? At the very least the proof that ($\Box \approx$) is valid is straightforward for s-rigidity. But in that case, an obstacle to Project 1 must be overcome. Skeptics who believe that terms have existential import (TEI) will deny that 'Hesperus' is s-rigid. The s-rigidity of 'Hesperus', they will say, commits us to the view that 'Hesperus' refers to Venus in all possible worlds. So if TEI holds, then Venus must exist in every possible world, and clearly there are possible worlds where Venus does not exist.

One way out is for the defender of Kripke to reject TEI. Evidence for thinking that Kripke actually *did* reject TEI is that variables in the QML presented in "Semantical Considerations" (Kripke, 1963, p. 87) are explicitly allowed to refer in a world to objects outside Ow, the set of things that exist in that world. On top of that, we now have direct confirmation from Kripke that he did indeed reject TEI and that s-rigidity has been his opinion all along about what rigidity should really amount to.[17] Mention of e-rigidity in *Naming and Necessity* was not intended as an endorsement, but was motivated by concerns to head off controversy which might blunt the force of his main message: that proper names and natural kind terms do not have descriptive content, but count as rigid in at least some sense.

A line of argument supporting the idea that it makes sense to reject TEI is related to Kripke's famous discussion of trans-world identity in *Naming and Necessity* (Kripke, 1972/1980, pp. 50ff.). There he takes pains to reject the metaphor of a possible world "as a foreign country". Accepting the metaphor prompts puzzles over how to evaluate 'Nixon might have lost the 1968 presidential election'. To decide this, we appear to need a necessary and sufficient description of Nixon so that we can locate which thing qualifies as Nixon in each possible world. Kripke's view on this is that no such description is needed. The referent of 'Nixon' is determined in the real world, so asking whether Nixon could have lost the 1968 election, amounts to asking whether *this man here in the real world*, has that feature in some counterfactual situation. Therefore rigid designation might be something of a misnomer—instead of saying that rigidity of 'Nixon' means that its referents are the same in each world, it would be safer to say that the referent of 'Nixon' is completely fixed to an object in the real world, and continues to refer to that object when we consider other counterfactual situations (possible worlds). But if this is how to understand reference in possible worlds, then reference should be insensitive to what exists in other worlds, and 'Nixon' would then refer to Nixon even when we are considering worlds (counterfactual situations) where Nixon does not exist.

I will leave it to the reader to determine whether Kripke or his defenders can marshal enough ammunition of this kind to convince the Skeptic to abandon the thesis that terms have existential import. In any case, this is a cost to be borne with the s-rigidity strategy. Not only that, but a second problem has to be solved. A fully developed QML containing terms that do not have existential import threatens standard rules for the quantifiers. Kripke's own QML (Kripke, 1963) had a language with variables only, and the classical theorem $\exists x x \approx y$ does not fail there because

[17] In oral comments on a presentation of a previous draft of this paper.

free variables are given the generality interpretation. On that reading, $\exists x x \approx y$ asserts the content of $\forall y \exists x x \approx y$, which is valid. However, it is not reasonable to think that terms like 'Hesperus' or 'water' can be given the generality interpretation. After all, 'Hesperus is bright' does not entail: 'Everything is bright'.

Therefore, if the demonstration of $(\Box \approx)$ is to be carried out in a QML at all, that QML must have symbols for proper names and natural kind terms other than the variables. But if we add any term t that lacks existential import to the logic, then the classical theorem $\exists x x \approx t$ will fail, presuming that the domain of $\exists$ is the set of objects that exist in a given possible world. That means that Kripke's project in "Semantical Considerations" (Kripke, 1963) to maintain classical quantifier rules cannot succeed here.

The cost may not be as great as Kripke might have imagined, because the correct QML for this case can be easily formulated using a strategy borrowed from free logic (Bencivenga, 1986)—a special rank of s-rigid terms is introduced, and the quantifier rules of Universal Generalization and Universal Instantiation are restricted to them. (See Garson, 2005). So while Kripke did not choose a QML based on free logic, that option is nicely suited for securing the argument that s-rigidity entails $(\Box \approx)$.

## 8.6 Will the Simplest QML Make a Difference?

It has been claimed that the only strategy for succeeding at Projects 1 and 2 is to presume that 'rigidity' means s-rigidity, and to adopt a QML based on free logic. Advocates for the simplest QML (SQML) (Linsky and Zalta, 1994; Williamson, 1998) may complain that this conclusion is too hasty. SQML preserves classical rules of quantification by assuming that quantifiers (and terms) range over a single domain O of quantification. Since $\exists x x \approx h$ is a classical theorem, the rule of Necessitation ensures that $\Box \exists x x \approx h$ (for Hesperus necessarily exists) is provable. As this is counterintuitive, an explanation is in order. It is that O contains more than actually existing objects, but also possibilia, so that what $\Box \exists x x \approx h$ claims is merely that it is necessary that Hesperus is a possible object. In this semantics, TEI (the principle that terms have existential import) is automatically satisfied as terms refer in O—the domain of quantification. So far, this paper has presumed, without comment, that the domains Ow for the different possible worlds w vary. The SQML approach is different enough that it is worth revisiting results of this paper from that point of view. The upshot will be that the problems raised in this paper are no better dealt with in SQML.

To establish that, let us review the three brands of rigidity in turn. First, consider e-rigidity, where a term t is e-rigid when t refers to the same object in all worlds where that object exists. If by 'exists' we mean: belongs to O, then the rider 'worlds where that object exists' is vacuous, since all objects exist in all worlds. That means that there would be no difference between simple rigidity and e-rigidity. However, Kripke's discussion in *Naming and Necessity* (Kripke, 1972/1980, p. 48), clearly distinguishes simple rigidity from rigidity. So the principle of charity demands that

we give 'exists' a different interpretation. The obvious alternative is to say that 'Hesperus exists' in a world w means that the possible object Hesperus is concrete, or actual, or realized in w. To express this formally in SQML, introduce a predicate letter C (for: 'is concrete'), whose extension v(w, C) varies from one possible world w to another. Given this machinery, it is easy to see that the result that ($\Box \approx$) is invalid for e-rigidity can be duplicated in SQML by replacing 'Ow' with 'v(w C)' in the definition of e-rigidity and the model M* defined in Sect. 8.2. So for the present discussion, the difference between a semantics with varying domains and SQML supplemented with C is merely notational.

A similar tactic may be used to show that the results on r-rigidity with stability semantics still hold. By replacing mention of the domain 'Ow' with 'v(w, C)' in the one or two places where 'Ow' is even mentioned in Sect. 8.3, it is easy to establish that ($\Box \approx$) is invalid given ($\Box$), and that the use of ($\Box$s) results in an interpretation of $\Box$ that is too weak to capture Kripke's doctrines about necessary identity.

Of the three brands of rigidity discussed, s-rigidity had the best prospects. Does the use of SQML in place of a system based on free logic help? It might seem so. The main objection to simple rigidity was that support for the view that proper names and natural kind terms were s-rigid entailed the need to reject TEI—the principle that terms refer only to objects in the domain of quantification. But it appears that SQML gets TEI for free, as it employs a single domain.

However, this "solution" is facile. The advocate of TEI is not concerned to argue that terms refer in a domain of possibilia. Her claim is that terms pick out only things in a world that concretely or actually exist there. That means that term h (for Hesperus) satisfies TEI only if it refers to an object in the extension of C (for concrete objects) in each world. Assuming that Venus is not concrete in at least one world, it follows that h is not s-rigid and there is little hope for completing Project 1. Therefore, the defender of Kripke cannot pretend to have solved the problem merely by adopting SQML.

However, the option of choosing s-rigidity and abandoning TEI remains, and SQML has the advantage over QML based on free logic of preserving the familiar classical quantifier rules. Even so, this does not, in my view, provide a decisive reason in favor of disregarding free logic systems. Within the language of SQML we will still want to be able to express such notions as actual or concrete existence, which can be done by defining the quantifier $\exists^C$x by:

$\exists^C \mathrm{x}\mathrm{A} =_{df} \exists \mathrm{x}(\mathrm{Cx} \,\&\, \mathrm{A})$

But the rules for $\exists^C$x are simply the rules of free logic, and so free logic emerges naturally within SQML in any case.[18] Furthermore, it is clear that the quantifier that matches Kripke's use of 'exists' in *Naming and Necessity* is not the possibilist quantifier $\exists$ of SQML, but rather $\exists^C$. Therefore, the best choice for formulating Kripke's

[18] More formally, free logic is embedded in SQML in the following sense. Consider the language containing only the $\exists^C$x quantifier (and its defined universal version $\forall^C$x) along with identity and the standard connectives. The provable arguments of SQML written in this language are exactly the arguments provable in the free logic system FL (Garson, 2013, Sect. 2.5).

reasoning is to use a free logic quantifier. Whether one finds it more convenient to define free logic quantifiers in SQML or to use a stand-alone system is in the present context a mere matter of taste.

I do not intend to argue here about the relative merits of classical and non-classical QMLs. As a logical pluralist, I see little value in insisting that only one option for formalizing QML must be right. In any case, that issue lies outside the scope of this paper. The takeaway point is that the adoption of SQML does not undermine the conclusion of this paper, namely that there are unexpected difficulties, and costs to be paid, in arguing from rigidity to Kripke's claims about necessary identity.

## 8.7 Conclusion

Three options for what Kripke might have meant by rigidity have been explored. Each of them faces problems in providing a link to support his doctrine that identities are necessary if true when they contain proper names or natural kind terms. It is surprising that the problems are so serious for e-rigidity, what many in the literature take to be the paradigmatic notion of rigidity—where a term is rigid when it refers to the same object in every world where that object exists. Unfortunately, this version of rigidity flatly fails to provide the needed argument for the validity of ($\Box \approx$), the conclusion that true necessary identities are necessary when their flanking terms are rigid. Furthermore, we now know that this version of rigidity is not the one that Kripke meant to endorse. Things are somewhat better for r-rigidity, where an r-rigid term refers to the same object in every world where it has a referent at all. At least on one version of stability semantics, ($\Box \approx$) is valid. But the costs are very high. The rules of modal logic need modification, and an unusual reading for $\Box$ must be chosen that trivializes Kripke's conclusions about necessary identities. The last option is simple rigidity. That choice requires that Kripke find a way to undermine the thesis that terms have existential import, for otherwise he cannot support his essential premise that proper names and natural kind terms are s-rigid. On top of that, if he introduces terms into the QML used to demonstrate that simple rigidity of terms t and s entails that $t \approx s$ is necessary if true, his project to preserve classical rules in QML will fail, presuming that the domain of the relevant quantifiers contains only existing (or concrete) objects. In sum, it is a lot harder to support Kripke's thesis about necessary identity than one might have first thought. Furthermore, the QMLs that have any hope of securing the right bond between rigidity and a robust conception of necessity are quite different from the one Kripke presented in "Semantical Considerations in Modal Logic".

**Acknowledgements** I am indebted to Cameron Buckner and to anonymous readers of a first draft of this paper for important improvements.

## References

Bencivenga, E. (1986). Free logics. In D. Gabbay & F. Guenthner (Eds.), *Handbook of philosophical logic* (Vol. 3, pp. 373–426). Dordrecht: Reidel.

Gabbay, D. (1976). *Investigations in modal and tense logics with applications to problems in philosophy and linguistics*. Dordrecht: Reidel.

Garson, J. (2005). Unifying quantified modal logic. *Journal of Philosophical Logic, 34*, 621–649.

Garson, J. (2013). *Modal logic for philosophers*. Cambridge: Cambridge University Press.

Gomez-Torrente, M. (2006). Rigidity and essentiality. *Mind, 15*, 227–259.

Hintikka, J. (1961). Modality and quantification. *Theoria, 27*, 119–128.

Kripke, S. A. (1963). Semantical considerations on modal logic. *Acta Philosophica Fennica, 16*, 83–94.

Kripke, S. A. (1972/1980). *Naming and necessity*. Cambridge: Harvard University Press.

Linsky, B., & Zalta, E. (1994). In defense of the simplest quantified modal logic. *Philosophical Perspectives (Logic and Language), 8*, 431–458.

Prior, A. (1957). *Time and modality*. Oxford: Oxford University Press.

Salmon, N. (1981). *Reference and essence*. Princeton: Princeton University Press.

Soames, S. (2002). *Beyond rigidity*. Oxford: Oxford University Press.

van Fraassen, B. (1969). Presuppositions, supervaluations, and free logic. In K. Lambert (Ed.), *The logical way of doing things* (pp. 67–91). New Haven: Yale University Press.

Williamson, T. (1998). Bare possibilia. *Erkenntnis, 48*, 257–273.

# Chapter 9
# A Letter from Kripke to Lewis

**Saul A. Kripke**

**Abstract** The following is a typeset copy of a letter sent by Saul Kripke to David Lewis on August 11, 1969 regarding the article "Counterpart Theory and Quantified Modal Logic" (Lewis, 1968). The original letter was typeset by Mimi Foster (indicated by the initials "mf" at the end of the letter) at Rockefeller University. In consultation with Saul Kripke, we corrected some typos, filled in blank formulas, and added three footnotes. Keywords have been added before the letter, references have been added after the letter, and the original pagination of the letter is reflected in bracketed parentheticals. (The Editors.)

**Keywords** Counterpart theory · David Lewis · Modal logic · Similarity · Translation

August 11, 1969

Professor David Lewis
Department of Philosophy
University of California
Los Angeles, California

Dear Dave,

I have just been reading and thinking about your paper on counterpart theory, and I wanted to send you some comments.

My own inclinations as to a way of speaking don't seem to be very much like yours, since I think it is literally Hubert himself and not a mere counterpart who would have defeated Nixon under certain circumstances. In the same way I think that I am literally the same person today who did something yesterday, contrary to what you say on the top of page 115. Maybe this isn't too important, since I can interpret you as speaking in the latter case of time-slices and in the former case, in

S. A. Kripke (✉)
The Saul Kripke Center, The Graduate Center, CUNY, Room 7118, 365 Fifth Avenue, New York, NY 10016, USA
e-mail: kripkecenter@gc.cuny.edu

© The Author(s), under exclusive license to Springer Nature Switzerland AG 2024

Y. Weiss and R. Birman (eds.), *Saul Kripke on Modal Logic*, Outstanding Contributions to Logic 30, https://doi.org/10.1007/978-3-031-57635-5_9

some sense, of "world-slices". Maybe you gain some generality over the other way of speaking, since, e.g., where in some cases I might have to say that the identity criteria over time for people and material objects are somewhat vague you can say that the counterpart relation is not an equivalence relation; so your method may have some advantages if one wishes to avoid vagueness. (For the modal case, however, see what I have to say in the formal comments below.) More important, perhaps, I don't think that *similarity*, in any significant sense, has much to do with individuation over possible worlds. It seems to me intuitively clear that I might have been a person radically different than I am now and that someone else at the same time might have resembled me in almost all the respects people would find important and that in that case we would not say that the other person would have been I. Anyway I won't elaborate on such philosophic questions, on which there might be disagreement, since I have some *formal* comments to make. The first of these is a quite conclusive objection to your theory the way it is set up now.

The objection I think you will accept as conclusive [p. 2] is this: Let $a$ be an actual individual with two distinct counterparts in some possible world. Then, under your translation of quantified modal logic, if the object $a$ is assigned as value to the variable $y$, $(x)\Vdash(x \neq y)$ comes out true, while $\Vdash(y \neq y)$ is of course false. So the schema $(y)((x)Fx \supset Fy)$ is not valid under your interpretation! The same counterexample works for Kaplan's interpretation, page 119, provided that $a$ has a counterpart in every world.

Has anyone pointed this out to you? I think the best way of fixing it would be simply to assume that an individual can have at most one counterpart in every world. An alternative might be to adopt the following more complicated translation scheme:

Given $\phi$, define $\phi^\beta$ by induction or the number of quantifiers and connectives in $\phi$. All clauses except those for $(\Box\phi)^\beta$ and $(\Vdash\phi)^\beta$ are the same as yours. For $(\Box\phi)^\beta$, define it by induction on the number of free variables in $\phi$. If $\phi$ has no or one free variable, the clause is the same as yours. If $\phi$ has two distinct free variables, $x$ and $y$, $(\Box\phi xy)^\beta$ is $x = y\&(\Box\phi xx)^\beta : \vee : x \neq y\&(\beta_1)(u)(v)(W\beta_1\&Iu\beta_1\&Iv\beta_1\&Cux\&Cvy \supset \phi^{\beta 1}uv)$.[1]

Here $(\Box\phi xx)^\beta$ has already been defined, since it contains only one free variable (it is assumed that if necessary the bound variables in $\phi$ have been changed so as to avoid confusion of bound variables in $\phi xx$). Also note that the second conjunct of the second disjunct is the way you would have translated $(\Box\phi xy)^\beta$, given $\phi^\beta$. The pattern persists. For three variables, translate $(\Box\phi xyz)^\beta$ as $x = y\&(\Box\phi xxz)^\beta . \vee . x = z\&(\Box\phi xyx)^\beta . \vee . y = z\&(\Box\phi xyy)^\beta . \vee . x \neq y\&y \neq z\&x \neq z\&\psi$,[2] where $\psi$ is the way you would have translated $(\Box\phi xyz)^\beta$, given $(\phi xyz)^\beta$. And so on for more variables. $(\Vdash\phi)^\beta$ is dual. Note that if each object in fact has at most one counterpart in each world, the translation collapses into the old one. Not only is

[1] The idea is that the second disjunct with $x \neq y$ should be written as Lewis wrote the original formula. [Added by SK.]

[2] This formula is poorly written. There should be a first disjunct where $x = y$ and $y = z$, and after the conjunction one should have $\Box\phi xx$. The second disjunct corresponding to the first disjunct as displayed should have $x = y$ and $y \neq z$, and $\Box\phi xxy$. Corresponding changes should be made in all the other disjuncts except the last, which remains as written. [Added by SK.]

the new translation much more complex than the old one, it seems to me dispensable. Where you say that an individual $a$ can have two counterparts, $b$ and $c$ in a single possible world, why not say that there [p. 3] are two distinct possible worlds, one in which $b$ is "the" counterpart of $a$ and one in which $c$ is "the" counterpart. In the obvious sense, the new translation allows counterparts to be used only one at a time; in that case allowing an individual to have two counterparts in a single world seems formally pointless. I think the following technical claim is true. (I haven't checked it in detail as is also the case with some other technical claims made below): The valid schemata in quantified modal logic under the new interpretation are just the same as those under the old interpretation with the requirement that each object has at most one counterpart in a world. (Perhaps the difficulties you got into by allowing an object to have multiple counterparts in a world constitute a plausibility argument against your definition of counterpart in terms of similarity, since given this definition the multiple counterparts do indeed seem plausible.)

Shouldn't you have mentioned that in addition to the controversial laws of modal logic like Brouwer's and Becker's principles that your interpretation rejects, it also rejects such ordinarily uncontroversial laws as $\Box(\phi \supset \psi) \supset .\Box\phi \supset \Box\psi$ and $\Box(\phi\&\psi) \supset \Box\phi$? This *isn't* meant to be a damning objection; what it means is that your $\Box$ operator behaves like the NMN of Prior's **Q** in *Time and Modality*. Probably you are already aware of this.

The following seems to me perhaps more serious. If a language contains a name "$a$" of an individual no doubt you would treat it semantically the way you treat a free variable assigned to value $a$. Suppose the name is eliminated in favor of a predicate $Ax$ (Quine's Ciceroizes). It would be natural to interpret the introduced predicate as holding in each world of precisely the counterparts of the object $a$. Then if $\phi a$ is replaced by $(\exists x)(\phi x\&Ax)$ for atomic $\phi$, $\Box\phi a$ (or, alternatively, $\Box\phi x$ where the free variable $x$ is assigned the object $a$ as value) need not have the same truth value as $\Box(\exists x)(\phi x\&Ax)$. If we translated $\phi a$ as $(x)(\phi x \supset Ax)$, the difficulty would arise for $\Box \sim \phi a$. This is really another form of the problem on page 119, second paragraph; but the following is perhaps more serious. Suppose $b$ is a counterpart of $a$, $c$ is a counterpart of $b$, $c$ is not a counterpart of $a$, and $\phi x$ holds when $x$ is assigned any [p. 4] counterpart of $a$ in any world but fails when $x$ is assigned $c$; then $\Box\Box\phi x$ is false when $x$ is assigned $a$ as value, but $\Box\Box(\exists x)(\phi x\&Ax)$ is true (provided $a$ has a counterpart in each world).

The difficulty of the preceding paragraph can be put another way. Your interpretation allows some laws to be valid for atomic formulae containing just one free variable, even though some substitution instances of them are not valid when the atomic formulae are replaced by formulae with extra free variables as parameters. How can you be sure that the monadic predicates in your formulae are not interpreted so as to refer to certain individuals not explicitly mentioned? (This happens with "Ciceroizes" and perhaps with more natural predicates as well.) You assume that your atomic predicates are, in some sense you don't define or explicate, "purely qualitative". I have some feeling for this, but am a bit queasy about the notion; anyway, this assumption certainly doesn't hold for all languages.

Finally, I am not sure that your counterpart relation really gives any additional generality over those who speak of identity across possible worlds or require that the counterpart relation be an equivalence relation. To consider one example: suppose we have objects $a_1$, $a_2$, $a_3$, in worlds $H_1$, $H_2$, $H_3$, where each individual is the counterpart of its predecessor, but $a_3$ is not the counterpart of $a_1$. Instead of speaking of an intransitivity of the counterpart relation, I think one could consider this a case of relative modality where the relation $R$ between worlds is not transitive. That is to say, we could have $H_1 R H_2$, $H_2 R H_3$, but not $H_1 R H_3$. The problem is, of course, that in the old model maybe $H_1$ *was* related to $H_3$ (especially in the non-relative case you consider where all worlds are related to each other). We can take care of this by postulating an additional world $H_3'$, qualitatively like $H_3$, and such that $H_1 R H_3'$, but $a_3$ (or the individual corresponding to it in $H_3'$), is not the counterpart of $a_1$ in $H_3'$ if in the old terminology $H$ was the counterpart of $a$ in $H_3$.[3] Now we have four worlds instead of three and the process may go on creating an infinity of additional worlds, but the counterpart relation can thus be made into an equivalence relation.

[p. 5] In other words, where you speak of an intransitive counterpart relation and a transitive accessibility relation, by posting extra worlds one can get the same effect by assuming the reverse. I hope this isn't too vague; anyway, it seems to me that the valid schemata of quantified modal logic are just the same if you allow an intransitive counterpart relation as they would be if you allow instead an intransitive accessibility relation. Of course there is a real formal difference in the fact that if the accessibility relation is transitive but the counterpart relation is not, $\Box\phi \supset \Box\Box\phi$ is valid for closed $\phi$, whereas under the alternative this would not be so. On neither alternative would Becker's principle hold if $\phi$ contains free variables. As I have said above, though, the distinctions your approach makes between closed and open formulae seem to me to be dubious.

I had no idea that this letter was going to be so long. Maybe some of my remarks are still too sketchy. I hope the first technical difficulty, about $(y)((x)Fx \supset Fy)$, which seems to me the most serious, is clear. Anyway, I want to ask you a favor. I hear fascinating rumors about your work on the counterfactual, which sounds very interesting. Could you tell me something about it? Regards to Steffi and to the people at U.C.L.A.

Best,

Saul Kripke

SK/mf

## References

Lewis, D. (1968). Counterpart theory and quantified modal logic. *Journal of Philosophy, 65*(5), 113–126.

Prior, A. N. (1957). *Time and modality*. Oxford: Clarendon Press.

[3] This clause makes no sense as written. I'm not sure what I meant at the time, there may have been a typing problem. [Added by SK.]

# Chapter 10
# Individual Concepts: Their Logic, Philosophy, and Some of Their Uses

**Saul A. Kripke**

**Abstract** This paper is an amended version of a talk Saul Kripke gave at the Eastern Division meeting of the APA in 1992 (an extended abstract was previously published as Kripke, 1992). It contains philosophical reflections and technical results concerning "Carnapian" quantified modal logic, that is, modal logic with quantification over individual concepts. The paper contains the fullest statement by the author available of (un)axiomatizability results he obtained in the 1970s. (The Editors.)

**Keywords** Carnap · Individual concepts · Quantified modal logic · Random variables · Unaxiomatizability

## 10.1 Editorial Introduction

In the standard model-theoretic approach to quantified modal logic—essentially that found already in Kripke (1959a)—individual variables are assigned elements of a given (nonempty) domain. Thus, variables stand for objects, in the usual sense of the term 'object'. There is, however, an older alternative approach, due to Carnap (1958), on which variables are assigned *individual concepts*, that is, functions from worlds to individuals (elements of a domain). On this approach, one might think of 'the third planet from the Sun' as being a kind of object in its own right, albeit one which is the Earth at our point in modal space, and other celestial bodies at other points in modal space.

The individual concept ("Carnapian") approach is closely connected with the contingency of "identity" thesis and, on a relatively straightforward formal implementation, it is not hard to see that $(x = y) \rightarrow \Box(x = y)$ can be made to fail.[1]

[1] See, for example, Hughes and Cresswell (1996, Chap. 18) for more details.

S. A. Kripke (✉)
The Saul Kripke Center, The Graduate Center, CUNY, Room 7118, 365 Fifth Avenue, New York, New York 10016, USA
e-mail: kripkecenter@gc.cuny.edu

© The Author(s), under exclusive license to Springer Nature Switzerland AG 2024

Y. Weiss and R. Birman (eds.), *Saul Kripke on Modal Logic*, Outstanding Contributions to Logic 30, https://doi.org/10.1007/978-3-031-57635-5_10

(Kripke was keen to stress that = is really *coincidence*, rather than *identity*, in the framework of individual concepts.) Informally, just imagine the domain is a 52-card deck, $x$ is the constant function picking out the Ace of Clubs, $y$ is the card at the top of the deck, and there are worlds for all shuffles of the deck; $x$ and $y$ coincide at some worlds, but not at all.

The history of this paper is as follows. Kripke was working on technical 'problems in quantified modal logic with individual concepts in 1975–1976.[2] On March 12, 1976, he reported the following results concerning **Q2**-style quantification over individual concepts with coincidence (=) in correspondence to David Kaplan and Richmond Thomason:[3] **S5-Q2** is axiomatizable (he provides an axiomatization), whereas **L-Q2** is not, for many standard base modal logics **L** (including **S4**, **T**, and **K**).[4] Kripke outlines the strategy of the proof of unaxiomatizability in the letter—propositionally quantified modal logics, antecedently known to be unaxiomatizable (Fine, 1970), are reduced to corresponding individual concept systems—but the argument is not carried out.

Kripke's results were not published at that time, but Kamp (1977) reports them. The results, that all individual concept systems contained in **S4.3-Q2** are unaxiomatizable while **S5-Q2** is axiomatizable, are attributed to Kripke by Garson (1984), where a proof is also given of the first result. In the last few years, Kripke did some work to reproduce his own proofs of his results, which he indicated do not (entirely) match those given in the literature, but he had not accomplished this by the time of his death.

Per a note of Stephen Webb's, this paper derives from an edited transcript produced by Webb from a recording done by Jeff Buechner of a lecture Kripke gave on December 30, 1992, at the Eastern Division meeting of the APA in Washington, D.C., between 1:30 and 4:30 p.m. (The *APA Proceedings* 66:2 indicated the talk would occur on the morning of December 29th, but the date was changed to the 30th in order to allow more time for the talk.) An extended abstract was previously published as Kripke (1992). As the abstract makes clear, the delivered paper is a combination of Kripke's technical results from the 1970s and (mostly) more recent reflections on their philosophical significance.

---

[2] In fact, he also worked on quantified modal logic with individual concepts considerably earlier (Kripke, 1992, p. 71), but the technical work described here was done in December 1975 and January 1976, as attested in personal correspondence; cf. Bacon (1980, p. 194).

[3] Regarding **Q2**, we suppose that Kripke had Thomason (1969, Sects. 10–12) in mind. The main relevant formal features, for Kripke's results, seem to be (1) that quantifiers range over absolutely *all* individual concepts and (2) that atomic predicates be treated extensionally (roughly, what is true of an individual concept at a world is just what is true of the individual that concept designates at that world, predicates being interpreted as usual). Note, in particular, that the unaxiomatizability results (to be stated) depend on quantification over all individual concepts; see Kripke (1992, p. 71) and Hughes and Cresswell (1996, Chap. 18).

[4] Thomason (1969, p. 131) reports that, "David Kaplan has informed me that there is no axiomatization of **Q2**—the set of **Q2**-valid formulas is not recursively enumerable. (This result is unpublished, and depends on work which appears in Kaplan's dissertation.)" See also Thomason (1970, p. 75, n. 1). Kaplan's proof, however, was apparently erroneous (Bacon, 1980, p. 194).

Kripke worked with the Editors on making changes to the transcript and elaborating on various points. He approved a complete draft for submission to refereeing in mid-August 2022. Besides appending this editorial introduction, we have made only essential modifications to that version (dated August 16, 2022) in producing this final version. Obvious grammatical and typographical errata have been corrected in the body of the text without notice. More substantial posthumous modifications made by the Editors have either been enclosed in double-angle brackets $\langle\langle \ldots \rangle\rangle$ (to be distinguished from Mr. Webb's editorial comments, which have also been retained, but using double-square brackets [[. . .]]; see Footnote 5) or made in footnotes attributed to the Editors. Most such substantial edits are ultimately due to the expert referees, Allen Hazen and Leon Horsten, to whom we are indebted for their extensive and valuable feedback. Thanks, as well, to Maria Aloni for some helpful comments.

The Editors

## 10.2 Opening Remarks

When the committee asked me to speak at this APA meeting, it was as part of a symposium with the title "Modal Logic."[5] I took this title quite literally and wished to give a paper that would have some actual work in modal logic in it—not, as might (I suppose) be imagined, philosophical discourse on various issues involving the foundations of modal logic, or the nature of possible worlds, or similar issues which have been much discussed in the literature. Of course I will make a few remarks about them in the course of the talk, and I may have a bit more than that to say about them, but the core of the paper is a set of theorems that I obtained some time ago.

This old work still left some open problems, and I had hoped perhaps to settle one or two of them by the time of this symposium, but either—and I am using an inclusive "or" here—either my research schedule or the difficulty of the problem mostly precluded this. There is one old result which I think I can now obtain in a simpler way than previously. Also, in a more recent period, I have been interested in the question of the extent to which individual concepts appear in various aspects of our thinking, and will try to get in a bit of that too.

I'd like to thank the chair, Billy Joe Lucas, very much for giving me an extended time. I had thought of myself as very rushed indeed, and I am still going to cut out at least one or two of the topics that would have been in the abstract. But at least I can explain a little bit more about what is going on here.

---

[5] [[*Note by Mr. Webb*. The process of transforming the tape into smooth text has made a certain degree of paraphrase inevitable. My degree of confidence that little substantive change has been introduced is high. But I have also produced a virtually literal transcript of the talk, and Mr. Buechner still has the tapes on file, if scholarship should for some reason require returning to the source. Wherever I write in my own voice, I enclose my words in double square brackets. Single square brackets are reserved for Kripke's use, for example in asides within direct quotations.]]

## 10.3 Individual Concepts

Now, trying in the worst way to avoid making a purely philosophical paper, I in fact chose a set of results that was motivated by a philosophical concern. The theorems pertain to an approach which is linked to, as many of you may very well know, a philosophical position that I have long opposed. However, my opposition to the position is to some degree with regard to its ill-conceived motivation, and the system is intelligible even without this. I think Ruth Marcus once attributed to Quine the view that modal logic was conceived in sin; I certainly think that this is the case with these systems whether or not it is true of modal logic *per se*. But I hope we are not so prejudiced as to think that bastards cannot grow up to be very fine, interesting people, and the same is perhaps true here. One of the things I like about these systems is the beauty of some of the theorems that one can obtain about them, and which are obtained in a way that I have found very aesthetically pleasing.

Now, Carnap (1958) should be regarded as the originator of this approach, although I have made no effort to follow him to the letter. To set this up, we assume a set of a usual kind for the modeling of modal logic: that is, we have a frame, or as I called it, a model structure $\langle G, K, R\rangle$, with a real world, a set of possible worlds, and a relation among the worlds. As usual, something is necessary in a given world iff it is true in all related worlds, and the semantics for the connectives are as usual.[6]

Take also a domain function $\Psi$ which, for these purposes, we will assume always gives a non-empty domain to each world. Of course philosophically this ought not to be required, and mathematically it is not necessary, but it makes the treatment of what I am doing here simpler. The system which I call "quantification over individuals" is standard as given in my papers or other papers following my approach. So when you have a formula

$$\Box A(x_1, \ldots, x_n) \tag{10.1}$$

with values given for the individual variables, we have a modeling $\phi$ which to every atomic predicate assigns an extension of an appropriate number of $n$-tuples in each world. If $A$, as it appears in (10.1), were atomic then (10.1) will be true of the individuals in a given world if and only if the model assigns that $n$-tuple of individuals into its extension. And in my way of setting this up the quantifiers ranged in each world only over each domain; in particular, notice that the predicate of identity is assigned true in each world if $x$ and $y$ are assigned the same value, and false otherwise. This automatically makes this formula

$$(x = y) \supset \Box(x = y) \tag{10.2}$$

come out true.

Now, there are really two nearly-contemporaneous and really rival approaches to quantifying into modal logic; the semantics thus given here agrees with Marcus's

[6] This last sentence occurs in the transcript but was inadvertently omitted from the typeset version Kripke reviewed. (The Editors.)

axiomatization, but disagrees with Carnap's semantics.[7] In the ensuing history, you see, many people regarded a formula such as (10.2) as obviously objectionable, and as implying that if George Bush is the president then that is a necessary truth, and perhaps therefore that he always will be, which might be used as an argument for invalidating the recent election results. Russell had already shown this to be wrong, but just the same people worried about the interpretation of $\Box Ax$ for this reason. Quine in particular made this famous, saying, for example, "necessary greaterness than 7 cannot be a property of a *number* $x$, because the statement that 9 is greater than 7 is necessary but the statement that the number of planets is greater than 7 is not necessary. But of which number, 9, that is, the number of planets, are you saying that it is necessarily greater than 7?" And you need to attribute the thing itself different properties as it occurs under different names. So Quine argued.

I'm not actually going to be talking about counter-arguments to Quine. I do presume that this argument is not a conclusive objection to this conception. However, a philosophical point that would come out of the modeling is built into this conception, viz., it was thought to be objectionable that one should have a system of quantification of this kind, because the proposition '$\Box Ax$' (with a value assigned to $x$) is to be interpreted as true iff $A$ holds of $x$ in all $\langle\langle$accessible$\rangle\rangle$ worlds. But that depends (so it is said) on the idea of reidentifying an object from world to world, and this was said to be objectionable.

Now let us call this position which objects to reidentification across worlds "anti-haecceitism," as it sometimes is called in the literature. More exactly, the view is as follows. Worlds are given in terms of objects and their properties purely qualitatively; two possible worlds that have the same arrangements of individuals, with the same cardinality and the same relations between them, might just as well be thought of as the same world; at any rate, there is no saying whether an individual from world to world is the same or different.

There is a standard way of expressing this sort of position mathematically. A model assigns to the possible worlds essentially an extensional first-order structure. If we have the modeling assigning extensions just to certain primitive predicates, and not necessarily to all but to those that are relevant to the formula we are considering, then we can consider the model as giving a directed system over the relation of ordinary first-order structures assigned to each element of $K$ (that is, to each world—a world can be thought of either as just the index element of $K$, or as the first-order structure assigned to the world itself saying, so to speak, the *way that it is* with respect to the predicates). And we can thus think of a modal modeling as a system directed by relation $R$ of ordinary first-order structures. Of course we know that in the case of **S5** any string of modal operators can be reduced to its last element, which for practical purposes means that we can dispense with this accessibility relation and say that a formula is necessary just in case it is true all over $K$.

---

[7] Actually, Marcus (then Ruth Barcan) gives two definitions of identity. Following *Principia Mathematica*, Barcan (1946) defines identity as second order indiscernability: $(F)(Fx \supset Fy)$. Then there is also a version where the conditional is strict: $(F)\Box(Fx \supset Fy)$. She proves them coextensive (in **S2**) and necessarily coextensive (in **S4**). Note that the second definition does not exactly express the necessity of the first.

Given that, the position of antihaecceitism is the position that identities between individual elements of the domains of worlds never matters; and the modal system should be set up in such a way that it never differentiates between two structures depending on whether an object is identified with another one (well, with the same one, I hope) in another world. The position can be expressed mathematically, and it standardly has been, by the idea that if you have an isomorphism between $K$ and another set $K_1$, and of $G$ onto $G_1$, and of $R$ onto $R_1$, that is, an isomorphism of the frames, and if you have corresponding elements of the frames the structures assigned onto them are isomorphic, then all modal formulae should have the same values in each structure. Notice that the isomorphism can be different in different worlds; so there isn't a global mapping of all of the elements of the unions of the worlds onto each other which then has to be fixed. That's what would have to be done for this one. We can do it separately for each world; and this means, intuitively speaking, that the truth values of formulae of the form $\Diamond Ax$ and $\Box Ax$ cannot depend on whether the same individual or a different one appears in given worlds.

This is very similar to a standard thing in mathematics: when people consider the direct product of a collection of rings, the direct product is thought to be completely determined as long as the factors are isomorphic, which are like the worlds; one does not have to have a global isomorphism mapping all elements in the unions of the factor structures onto each other.

The system of quantification over individual concepts was conceived, although not explicitly (I must say) in this form, as one means of getting an antihaecceitist system. Following Carnap, we take an individual concept to be a function mapping elements of the domain $K$ to individuals. We will require in this case that the individuals all be elements of the domain of the worlds in question. So it's a map mapping any world to an element of the domain.[8]

(The values assigned to the variables are individual concepts. Since atomic formulae are given just as extensions of each world, they can be interpreted for a given system ⟨⟨~~the values to~~ of values of⟩⟩ the variables—I mean, if $A$ were atomic, ⟨⟨for example, if⟩⟩ it's $Ab$—iff the extension assigned to $A$ holds of the $n$-tuple of individuals assigned to $b$ in a given world. This makes the atomic formulae purely extensional in the sense that they don't depend on the individual concept that is involved, but only on its value.)

We could consider a system of quantification over individual concepts that does not have this restriction. Given this restriction, why do the individual concepts matter at all? Maybe only their extensions matter. Well, intensionality arises when we have modal operators applied to formulae containing individual concepts. So, in particular, the identity predicate in (10.2) is interpreted in such a way that for two given individual concepts, ⟨⟨$x$ and $y$⟩⟩, to say "$x = y$" is not really to say that they are the same individual concept (which could be written $xIy$); it is to say that their extensions coincide. The formula (10.2) will therefore be false.

[8] In some contrast to Carnap, we choose not to take definite descriptions as primitive, but to use Russell's theory.

Sometimes in the philosophical literature, this type of argument has been taken as an argument against the necessity of identity, but it really shouldn't be. It's just that predicates that originally are predicates on individuals induce corresponding extensional predicates on individual concepts, in the sense of individual concept we have been laying out. And a peculiarity of this is that (10.2) is false for the latter sort of predicates.

I should mention that in this setup I will take the individual concepts assigned as values to the variables to be *total*; that is, they have a value at each world, and that value is always an element of the domain of the world. I've done this not because it is philosophically the best way of looking at things, but for mathematical convenience, and with the fairly confident thought that the more general case allowing partial concepts will, mathematically speaking, reduce to this special case—at least for most of my present purposes.

I also am mostly considering the system where the system of individual concepts comes from a basic underlying one of individuals, so to speak, so that the atomic predicates are all induced by predicates of individuals. We could start out with intensional concepts from the beginning; then the intensionality would come in on the atomic level. But mathematically this case is much less interesting because it's much more trivial. Also, a good deal of the motivation of what Carnap was doing is lost without this restriction, or without at any rate some loosening of the restriction, but not a full loosening of it to arbitrary intensional concepts. Even Carnap did say things which would imply that some loosening of the restriction could be involved.

I want to say a couple of things about this. First, I've used the terms "individuals" and "individual concepts" because this is customary. But in a way, it seems to me that labelling these rival systems in this way is of a misleading character. First, the individual concept system is obviously intelligible to any modal logician, regardless of whether the modal logician accepts the motivation in which it was conceived—in sin, supposedly. It is certainly intelligible to me. The questions of whether it is useful and interesting are something else; and I did indeed find it interesting, and some extensions of it to be useful.

Often in the literature these two systems have even been regarded as rival forms of quantification over individuals. In that sense I think the advocates of the Carnap-type system were certainly wrong. When we say someone who lost an election might have won it, as, for example, some people in the White House are saying that they might have won this last election if it weren't for Walsh, we are referring to a particular which we identify across worlds. However, I should say the term "individual concepts" has been rather misleading insofar as it suggests that a particular metaphysical status, say that of concrete particulars of some kind, is being stipulated into the values of the individual variables. All other forms of extensional quantification theory have been said to be over any domain whatsoever, and I don't see that this point of view should be abandoned just in modal logic. So what I have been calling quantification over individuals is really quantification over entities. In fact, these entities could be the individual concepts themselves. The standard system as I was describing it applies to any entities where one feels one can speak of the modal properties of the thing itself in another counterfactual situation. So for an individual concepts system to be a

rival form of quantification, it would need to regard quantification over individuals as unintelligible across the board. It is probable that there is no coherent philosophical position which does this and also accepts modality in any way.

In particular, if someone believes in the antihaecceitist position with respect to some particular domain of objects, say concrete particulars, and accepts the local isomorphism criterion for the intelligibility of the system, then such a person should regard quantification over individual concepts as extensional. And the ordinary system, with its domain being given that of individual concepts, can be directly interpreted as a system over individual concepts, where, however, identity will be identity of individual concepts.

Quine claimed at two different stages of his life that if you did make a restriction to individual concepts his strictures against quantified modal logic would disappear, but that this involves repudiating the entities of the concrete world; he later went back on this position and said that, on the contrary, his strictures were valid against every sort of quantification, if valid against any sort at all. The justice of the latter remark may depend in part on what you take his strictures to be, and how much of a fallacy one is going to allow to be included in the antecedent, "if my strictures are valid at all." However, I disagree with any position that would say that even from an antihaecceitist point of view, quantification over individual concepts—with identity interpreted as identity of individual concepts, and with the primitive predicates applying to individual concepts—would be unintelligible because it involved essentialism.

Of course it is true that for any domain of objects you can find definite descriptions that pick them out the same way in the actual world but differently in other possible worlds. But this is hardly relevant if the position holds that only some descriptions do pick out the same object in all worlds, or perhaps that none do, because this means that the position itself insists that descriptions (or at least the non-constant ones) are irrelevant. This is the case here, because once we have the antihaecceitist system $\langle G, K, R\rangle$ over some domain, the haecceitist system with individual concepts is perfectly intelligible with only local isomorphism, and what an individual concept is across worlds is clear: they're not even in worlds, they're given by functions from these worlds to objects defined in a particularly global (so to speak) way. So this remark by Quine is to be rejected.

On the other hand, something is still open to us to make peace with the antihaecceitist. I partly got interested in this system from following my belief that the modal logician should not necessarily just stick to his or her own philosophical position, but should explore the formal consequences of different philosophical positions. But I have wondered: is there a domain of entities such that the antihaecceitist system is the plausible thing for that type of domain? I think there are probably two cases of this. One is: descriptions of possibilities that are by themselves given purely qualitatively. An example given in Feller's probability textbook is the distribution of misprints in a book, and the statistics that is relevant (Feller, 1968, p. 42). And the second is: there is a case for a certain amount of antihaecceitism in quantum mechanics. I don't know how much of a case; the situation is somewhat unclear to me, mostly because quantum mechanics and its foundations are so unclear that I have hardly thought about

it in a serious way. Certainly I am not one of the experts. Anyway, this situation in quantum mechanics is mentioned by Quine (1976). Quine's later views don't include ordinary quantification over physical objects, but just over points of space-time, and physical objects are occupied regions of space-time; they really come out as certain types of predication. In "Whither Physical Objects," he says:

> The naive conception of matter [the one he is getting rid of in this paper] thus tends to dissolve in modern physics. In fact its dissolution has gone yet farther, to judge from an account by Heinz Post. Not only can it be arbitrary to identify an electron over time; there seems even to be something not quite right about distinguishing between coexisting electrons in space. (Quine, 1976, p. 498)

Thus consider two boxes, which I shall call the east box and the west box, and two electrons, $x$ and $y$. Common sense recognizes four ways in which $x$ and $y$ could occupy the boxes: they could both be in the east box, both be in the west box, or one in each for the other two ways.

| EAST | WEST |
|---|---|
| $x, y$ | |
| | $x, y$ |
| $x$ | $y$ |
| $y$ | $x$ |

We have four possible distributions, and common sense would say that they should each be ascribed a probability of $1/4$. But according to Post, these last two apparent possibilities have to be counted as just one: the fact of $x$ being in the east box and $y$ in the west box, and the fact of $y$ being in the east and $x$ in the west, must be counted as just a single possibility, contrary (Quine says) to common sense. The moral of this is to not think of individual electrons $x$ and $y$ at all, but of states of the boxes; states of being singly occupied or doubly occupied are innocently all states of being singly or doubly affected. The east box may be doubly affected, the west box may be doubly affected, or both boxes may be singly affected, and there the matter ends. There is no question of which electrons at all as entities, only of singly or doubly occupied boxes.

Actually this passage is not, I believe, correct. (Is there an expert here on this?) Electrons obey the so-called 'Fermi-Dirac' statistics, which excludes there being two in the same box. But there are other particles—photons, I think, are examples—which have the appropriate spin and do not exclude one another. And such particles obey the 'Bose-Einstein' statistics. But at any rate, there certainly are particles for which this description by Quine is the correct distribution. Both distributions are anti-haecceitist.

Quine's first response is: "this is very surprising; wouldn't you think that there are four distinct possible worlds?" Well, people might, but that just shows that people don't think enough. The real analogue to modal logic is probability theory; tense logic and time is a certain analogue, and other things with indices, but they miss the sort of modal dimension about being counterfactual possibilities. But probability theory is a real analogue to modal logic. And notice that Quine here thinks it very

strange that one should think of only three distinguishable states rather than four. In this case, indeed, Quine's position does agree with common sense, no doubt because he is not thinking of this as being about modal logic.

If you did think about things in the intuitive way and insisted that there were four distinguishable states, and then used a philosophical indifference principle to assign equal probabilities to each possible state of the world, you would get the 'Maxwell-Boltzmann' statistics. But the surprising thing is that elementary particles as they are to be found in contemporary physics do not obey these; instead, something of an antihaecceitist sort is obeyed by these particles. These two possibilities are distinct, but these two are identified as being the same qualitative distribution.

There's something a little otherwise funny about this too, because here the distributions might be thought to be distinct as to which was the real and which a possible world, or something like that. But these are unknown worlds, where nothing is known to be real. You could get a setup with three or four particles; they will still be indistinguishable, even if a particular world is picked out to be real. So something is rather like a sort of antihaecceitist probability about this, and far from regarding this as a vindication of our naive intuitions, this was regarded by physicists and by normal lay readers such as Quine as one of the paradoxes and surprising results in quantum mechanics—not something we would have intuitively thought to be the case. Nevertheless, it does show that for certain purposes in quantum mechanics, an antihaecceitist reading of the quantifiers and of the entities involved is perhaps an appropriate one.

It is also true, I perhaps should say, that the Carnap system is not the only antihaecceitist approach; there are many others, including those that quantify directly over individuals. Indeed, you can get an antihaecceitist interpretation of quantification over ordinary individuals into modal contexts in this system. There are a couple of them.[9] If I have a chance, I will try to say what this is; it is fairly easy. On the other hand, in quantum mechanics and in all other probability statements, I don't think we have a pure antihaecceitist setup. Let us consider, for example, the following well-known type of situation. We have a detector plate and a photon which is supposed to hit the plate, though it may not; it can only be predicted to hit the plate with a certain probability. We actually send many photons, a stream of them one after another. Let's say each of them has a probability of $1/2$ of hitting the plate. However, we send many, many photons, so there is a probability of nearly 1—say, .999—that at least one will hit the plate.

Suppose, by luck, the very first photon we shoot hits the plate. The naive physicist will say: this photon had a probability of only $1/2$ of hitting the plate, and it in fact did so. However, a Quinean physicist could and should say: no, having a certain probability of hitting the plate is not a property of a photon $p$; if you described this photon as Sam the Photon, perhaps it did have a probability of $1/2$ of hitting the plate. But let us describe it as the very first photon to hit the plate, or the very last photon to be shot if none does. (That's an individual concept.) Then that will hit the plate as long as any photon hits the plate, so its probability of hitting the plate will

[9] A referee suggested consulting Kripke (1985) on this point. (The Editors.)

be .999. But which object, Sam the Photon or the first photon to hit the plate, were you speaking of? It is not the photon itself that has a certain probability of hitting the plate, but only a photon under a certain description. So the Quinean physicist should say. Now, this argument is unlikely to be made in physics. I thought of sending it in to a journal so I could make my name in this area, but after further consideration thought the paper was not too likely of acceptance.

Of course, with ordinary objects such as dice, concerning questions such as the probability that a given pair of dice will fall in a certain way on the first throw when it is going to be thrown several times, you could make exactly the same argument and it would be equally fallacious. This shows, I think, that the way we think about probabilities of events happening to things is not in general antihaecceitist, and is not even so in physics. However, and I haven't thought enough about the matter ⟨⟨~~to do so~~⟩⟩, there is a case for thinking that certain restricted physical situations are somewhat antihaecceitist in character. So one doesn't need to favor an all-or-nothing position called haecceitism or antihaecceitism. "Haecceitism" is also too simple a term anyway; but I won't go into that, since I don't know what my exact views are. But as I have said, I do think that talking about possibilities about things is intelligible, and this would seem to endorse the use of a haecceitist account for some purposes. In general, what one's position is should depend on the domain involved, and even on the types of predicates to be applied to the elements of the domain.

I should perhaps say one more thing about this, but I'll have to be very sketchy since I can only give the topic a couple of sentences. One reason the individual concept system has been viewed as a system of quantifying over individuals is what at least one philosopher has called the "Fregean interpretation" of the quantifiers.[10] The idea of this interpretation is simple: take an existentially quantified statement,

$$(\exists x)Ax \tag{10.3}$$

where no free variables are involved other than $x$. The standard interpretation of this (Tarski's) takes as a basic notion the satisfaction of a formula, or as its truth when certain values are assigned to the variables. An alternative way of looking at things is to say that if some name for some object $a$ in the domain (I'm just thinking of an ordinary object) were added, then $Aa$ would be true. Now, in ordinary extensional logic, these approaches are pretty trivially equivalent. But in intensional logic one has the following choice. If one wishes to adopt this approach, one ⟨⟨~~either has to~~ can⟩⟩ say that certain constants are genuine names, and therefore are modally rigid, and only those can be substituted. It is because of this way of looking at it, I think, that the whole debate over Quine's position became so prominent: though in every other instance he wishes to eliminate singular terms, here very often in his exposition he looked at quantification in this way, writing as if the chief question was what privileged class of singular terms should be granted the power of substitutability.

Now, suppose that one thought that there was no privilege: that such a term as "the number of planets" could be put in just as well as any other, and that the formula

[10] See Evans (1985, pp. 80–87).

should be read naively, with no thought to scope or assuming minimum scope. If one doesn't allow vacuous individual concepts, that is, individual concepts that are the senses of vacuous descriptions, one will get from this idea precisely the Carnap-type system, and it might look from that point of view as if it were the natural extension of ordinary first-order quantification. Given those premises, it is natural, and in particular, one doesn't have to make a fuss about how identity really means coincidence of concepts here, since equals is just read as equals, and individual concepts don't explicitly come in. I think this is the way a lot of people did look at it, and it is an alternative way of looking at the system of quantification over individual concepts. With more time, I would say a bit more about this.

Let me say something else about Carnap's interpretation of his system. Carnap interpreted the system in a way that involves what I will call "Carnapian doublethink." This term is meant to be partly pejorative, but it also involves a certain degree of praise. Carnap said every term had an extension and an intension. In practice what this meant was that he observed that when no modal operators are present in a formula it is purely extensional, and it really doesn't matter whether we interpret the quantifiers as ranging over individual concepts or over individuals. Only the individuals (that is to say, only the extensions) are relevant, and it can be interpreted as ordinary quantification. ("Individuals" here means nothing metaphysical—it just means the elements of the domain, whatever that might be.) When modal operators are present, the intensions are relevant, and are used.

Quine criticized this very severely, and said "look: whatever you say, your system does quantify over intensions rather than over extensions." It is difficult not to sympathize with Quine here, at least in part. Nevertheless Quine goes a bit too far. First he says that the ordinary interpretation of the most ordinary formula will be hard, because the individuals in the concrete world have disappeared. He gives examples: taking numbers to be individuals, we say 8 is a power of 2; and he says, we write this out in just an ordinary way: we say, $(\exists z)(8 = 2^z)$; but this formalization depends on the customary interpretation with numbers. Here you would have some strange predicates of concepts, and it is hard to even know how to interpret the system. Another example he gives is "the wives of two of the directors are dead."

This of course is quite false, because (as Carnap said) one formalizes things like this in quite the ordinary way, with the ordinary predicates, the point being that in Quine's own terminology, coextensive individual concepts are indiscriminable in these contexts. The strange thing is that later on Quine thought that identity meant indiscriminability with respect to all the predicates. He thought that when you introduced a new predicate, you in effect reopened all identity questions.

Precisely the same point of view can be taken here in the narrow language, before modal operators are introduced: coextensive individual concepts might just as well be thought of as identical. It is only in the richer language that we have to think of these extra distinctions. Otherwise Quine would have to criticize his own apparatus and say that when we introduce a new predicate we are repudiating our own ontology, we make all of our past formalizations come into question, and so forth. And that is not true.

I should add that probability theory itself is rich in Carnapian doublethink. The doublethink in question is, for example, when one asks: "what is the probability that this pair of dice will fall a 9?" One is asking an intensional question here about the individual concept here. And then one says, "Oh, gosh, the dice *did* fall 9!" One is thinking about a particular matter of fact, and probably just thinking about the dice themselves. If one thought of this as one question, one is really using a sort of Carnapian doublethink here.

If you look at the textbooks of probability theory, they have this sort of doublethink. I mean not just that they use this sort of doublethink, though that is true too, but that they use individual concepts very explicitly and with the same formalization as in modal logic. The origin of the individual concept idea in Frege's notion of sense gives it an air that's associated with a very particular philosophical position. At one point it made Quine contemplate the idea that you couldn't have these things existing simultaneously with ordinary things, that they drive the ordinary entities out, because they are such spooky entities. But in fact these entities were, before they ever came along in modal logic, the bread and butter of probability theory. They were called *random variables*. The probabilist considers the space $S$ of possible states of the world and the probability distribution on these states. A random variable is a function from these states to objects. Examples given in probability texts are, for example, the case of the dice that I already gave, the number of aces in a hand of bridge, the number of multiple birthdays in a company of $n$ people, that is, the number of birthdays that occur more than once: the probability of that might be considered, the number of successful runs in $n$ Bernoulli trials—$n$ Bernoulli trials just means running the same gambling game again and again, for example, you roll the dice a number of times; what you get is the number of times the gambler wins, for a fixed number of runs. These are also events that can come out with certain probabilities. And they are formalized by the probability theorist by a function from states to, in these cases, numbers, and then in terms of the ground probability distribution the probabilities of the values is given.

Where $y$ is a random variable you can ask what the probability is of its having (say) a value strictly between 0 and 1. Here, the random variable just is a Carnap intension just as in modal logic; only it has probabilities attached to it. So, far from this being something introduced by philosophers, say along with a claim that it might interpret natural language, these entities have been around for some time. I don't know whether it was Kolmogorov who introduced them; they are at any rate a staple of probability theory. The difference between a random variable and an individual concept is simply that a random variable has a probability measure attached to it, and restrictions on the function so the probabilities in question make sense. There is also really in effect a nonstandard modal operator, where necessity means "holds almost everywhere" or "certainty". And often, if one says "$\Box(y = z)$", one can think of that as identifying the two individual concepts (or random variables) in question. But individual concepts have been very common in this branch of mathematics.

I don't think it is really controversial at all that there is such a link between individual concepts and random variables, although it probably has never been said in print. As I say in my abstract, Martin-Löf once asked me whether this has been

noticed, and I said I didn't know whether it had in general been noticed, but that I had often mentioned it in classes.

Let me mention another case where I think individual concepts are involved. I shall then get to my actual theorems. This case might be more controversial, and it's a case in which I think individual concepts are widely involved in science. So the use of individual concepts would then be even more pervasive than the above would indicate. The case is elementary calculus and analysis; and the link to science I especially have in mind is that calculus/analysis is used in physics. I should note that all of the basic features of this individual concept way of treating calculus have been pointed out by Kit Fine.[11] So far as I know, he's only given short accounts in print of his treatment of the calculus—I have heard a more elaborate account in a talk.

Individual concepts arise in calculus as follows. The ordinary way of writing a derivative is as $\frac{dy}{dx}$. So from the start the derivative looks as though it has two parts, perhaps related as numerator and denominator. And this is to some degree borne out by ordinary usage. Suppose we have an equation such as $z = \sin x^2$. Then to compute the derivative of this, one might set $y = x^2$, and therefore $z = \sin y$. And then one computes $\frac{dz}{dy} \bullet \frac{dy}{dx}$; that is, we take the product of these two derivatives. The derivative of $z = \sin y$ will just be $\cos y$; and the derivative of $y = x^2$ is $2x$. So the product is $\cos y \bullet 2x$; and since $y = x^2$, the derivative of $\sin x^2$ comes out $\cos x^2 \bullet 2x$. And in doing all of this, one is appealing to the Chain Rule:

$$\frac{dz}{dy} \bullet \frac{dy}{dx} = \frac{dz}{dx} \tag{10.4}$$

In words: the derivative of $z$ with respect to $y$ times the derivative of $y$ with respect to $x$ equals the derivative of $z$ with respect to $x$. Each derivative would seem to consist of two entities rather than one here. And the '$d$'s, which give the Leibnizian idea of the differential in infinitesimals, can really be read as just notational, so that $\frac{dy}{dx}$ is simply a funny notation for $d(\frac{y}{x})$, or better yet $d(y, x)$. And so a derivative seems to be an entity which is determined by two *variables* of the ordinary sort. This way of thinking about derivatives, and this form of the Chain Rule, is the way it is taught to elementary students.[12]

But in the conventional formalization this is a fiction. The conventional notion of the derivative takes the derivative to be an operation on a function $f$. It has one

[11] For some of Fine's brief remarks (using his more complicated arbitrary objects), see Fine (1985, pp. 142–3). Philosophically, I might be somewhat critical of Fine's approach. He seems to identify two distinct notions: a fixed, but arbitrary (or indeterminate) object, and a variable quantity. Perhaps some such ambiguity might apply to the interpretation of individual concepts. But in my present discussion of elementary calculus and analysis, I intend the 'variable quantity' interpretation.

[12] It should be stated that there is a modern theory of differential forms, applying to multivariate calculus, that gives an interpretation of differentials, so that if $y = f(x)$, $dy = f'(x)dx$ holds as a theorem. And similarly for the multivariate case with partial derivatives. This theory, which might be regarded as somewhat formalistic, will not be explored here.

argument, and its value is another function $f'$. This means that the basic equation involved here will be something like the following, where $\bullet$ is multiplication and $\circ$ is composition:[13]

$$(f \circ g)' = (f' \circ g) \bullet g' \tag{10.5}$$

This is the way the Chain Rule is expressed in the official functional notation. You see that it is harder to memorize than anything like (10.4). In a talk, Kit Fine went through the formulation and proof of this, including the proof of the chain rule, in conventional function notation, and showed how messy it was compared to the notation used in (10.4).

Fine, by the way, attributed this conventional method of formalization to Frege, and even writes as if the mathematicians don't use it. But that's not right; the mathematicians do use it. They've never heard of or read Frege, but this functional notation is the way they think derivatives should be represented for theoretical mathematics. And in practice they're more and more likely to use the functional notation as the mathematics gets more rigorous.[14] Now, Kit Fine viewed this in terms of what he called his theory of arbitrary objects; but I wish to do it in terms of individual concepts, and I will say it as follows.

In general when one has an equation relating physical quantities, such as

$$y = \frac{1}{x} \tag{10.6}$$

or

$$y = x^2, \tag{10.7}$$

these are the values of the variables that may vary over time or over the possible situation. For example, if at constant temperature the pressure is inversely proportional to the volume, this means that both over time (in a limited situation) and over counterfactual states—the modal dimension is there too—one quantity is a function of the other, or they have some systematic relation. It is the distribution of these values that is relevant; and what I would propose here is that one has a case of individual concepts, together with the appropriate Carnapian doublethink at times, where one thinks of the variables as standing for the relevant individuals, or as given values of the individual variables, when the modal properties do not matter. This type of doublethink is actually common in the calculus textbooks, and it could lead to the question "what are $y$ and $x$—are these just numbers?" That is why it is said that, rigorously speaking, we have to represent these as functions and use lambda notation and so on. But this is an alternative approach.

I only have but so much time that I can spend on this, but I want to say that what is interesting here is that not only can these entities be understood as indi-

[13] In the transcript, (10.5) was $(f \circ g)' = (f' \bullet (f \circ g'))$, which we have replaced with the correct statement of the Chain Rule. (The Editors.)

[14] Engineering students, or even physics students who are less concerned with mathematical rigor, may be less likely to be taught (or even to care about) the 'rigorous' functional formulation.

vidual concepts, but also the modal language, enriched by relevant mathematical concepts $\langle\langle$~~and one other thing~~$\rangle\rangle$, is highly suited for expressing the relevant elementary mathematical concepts. So this is an alternative way of looking at things, both to the conventional mathematical way and to Kit Fine's proposal to use his arbitrary objects, which can in fact be viewed as a special case of my approach.

If we have an index set of states of the world, or instants of time, or both, of such pairs of things, under which (10.7) is supposed to be an individual concept giving a physical quantity, say, which is a real number, then such things as $y^2$ and $y^3$ are clear individual concepts. One form of Carnapian doublethink is going from the "is" of simple identity between individuals or coextensiveness of the concepts to meaning-intensional identity. Here we take an **S5**-type necessity, and to equate $y$ and $x^2$ is to write (if "=" is the extensional equals) "$\Box(y = x^2)$". Or we could have the primitive "$I$" (for identity of individual concepts); it $\langle\langle$(i.e., $I(y, x^2)$)$\rangle\rangle$ means that $y$ is $x^2$ at all points. In general we can say that under certain circumstances $y$ is a function of $x$. Here it is useful, and certainly is done in calculus, not to succumb so much to the strictures of antihaecceitism. Certain individual concepts are *constants*: they take the same values in all worlds (or at all times). They are usually represented in calculus by special letters. We might as well identify every real number—or complex number, or whatever it is—with the corresponding constant that constantly takes its value. So we will do so. In general, $x$, $y$, and $z$ will be over individual concepts, other letters of the alphabet over constants. So "$y$ is a function of $x$" will be expressed as:

$$\Box(r)(\exists s)\Box(x = r \supset y = s) \tag{10.8}$$

In other words, for every constant $r$ there will be a constant $s$ such that it is true at every index that if $x$ takes the value $r$, then $y$ takes the value $s$. Here $r$ and $s$ are constants; and of course $y$ can't take two values, that's built into the thing.

There is also continuity and the derivative, for example, to express. These too are nicely expressed in the modal language. It might be nice sometime to trick this out with things like double indexing and so on, but in general I think I'm going to avoid that at this level. So what I mean in saying that individual concepts are appropriate is not only that these entities are appropriately used, but that the modal language is the appropriate way to express things.

The property of continuity can be expressed as:[15]

$$\begin{aligned}(x)(\epsilon > 0)(\exists\delta > 0)(r)(s)((x = r \wedge y = s)\\ \supset \Box(|x - r| < \delta \supset |y - s| < \epsilon))\end{aligned} \tag{10.9}$$

Now "$|x - r|$" here is what is called $\Delta x$ in the calculus, where this is the difference between the actual value of $x$ and the value it would have had somewhere else. I'm

[15] In the transcript, (10.9) was $(\epsilon > 0)(\exists\delta > 0)(r)(s)((x = r \wedge y = s) \supset \Box(|x - r| < \epsilon \supset |y - s| < \delta))$, which we have replaced with the correct epsilon-delta statement of continuity. (The Editors.)

using orthodox calculus here, not nonstandard analysis; there'd be some way of using individual constants to understand that.

Traditionally, in calculus classes, one expresses continuity and especially the definition of the derivative in terms of quantities called $\Delta x$ and $\Delta y$. They look as if they are functions of entities $x$ and $y$, and mean the same thing when applied to both $x$ and $y$. However, in fact, in the traditional formalism, where $y = f(x)$ holds in a suitable neighborhood, $\Delta x$ is an arbitrary quantity $h$ such that $h \neq 0$ and $\Delta y$ is $f(x - h) - f(x)$. In the present approach, we assume that there is a focus point (representing the present time or actual world), and $\Delta x$ and $\Delta y$ represent derivatives of $x$ and $y$ from the focus point. That $y$ is a continuous function of $x$ at the focus point can be expressed in terms of $\Delta x$ and $\Delta y$; and the derivative can be expressed in traditional terms as $\lim_{\Delta x \to 0} \frac{\Delta y}{\Delta x}$, where it must be assumed that $\Delta x$ has non-zero values in an arbitrary neighborhood of the focus point.

For the chain rule there are, just as in the traditional formulation, two cases. In one, the degenerate case, $y$ is constant in a neighborhood of the focus point, so that the derivative at the focus point is 0. Then if $z$ is a function of $y$, $z$ is also constant in a neighborhood of the focus point, and the chain rule holds. Otherwise, $\Delta y$ has some non-zero values in every neighborhood of the focus point, and $\frac{\Delta z}{\Delta y}$ is defined for arbitrarily small neighborhoods of the focus point. The traditional argument for the chain rule in terms of $\frac{\Delta z}{\Delta y} \cdot \frac{\Delta y}{\Delta x} = \frac{\Delta z}{\Delta x}$ can be carried out. If one contrasts this with the traditional exposition in terms of composition of functions, the greater intuitiveness of this formulation should be clear.

This individual concepts understanding of calculus seems to me to be intuitively what is going on when one deals with physical quantities, because one intuitively wants to say both that a physical quantity is a quantity and yet that it indeed varies over time. There is still some other work to be done on it; for example, I wanted to see what the most natural interpretation, if there is one, of the '$d$'s would be in terms of this approach.

Fine's approach is very much like this one. His approach really involves structured individual concepts, which are structured entities which need not be identical even when they coincide over all worlds. And a case could be made here for using these structured individual concepts sometimes. But one point where Fine disagrees with standard mathematical usage is that in using calculus we sometimes speak of two variables, either one of which could be taken to be the independent variable and the other the dependent variable; and his approach does not allow this. My version certainly does allow this; you just have two individual concepts, and either one of them may be a function of the other. Or you have $n$ individual concepts; there could be a relation between them which determines implicit functions, and so on.

So I hope you get the flavor of this. Some of this has yet to be worked out; it isn't as though I've been working on it again in this last month, except in getting ready for this talk, but I worked this out quite a bit more recently than the theorems in question.

I think also that in some things in natural language where we speak of such relations, if we think of entities involved, they are individual concepts. I should point out something already observed by Frege and Russell, or really by the joint work of

them. Frege observed that when you have a statement that might appear to involve a definite description about ordinary individuals inside an intensional operator, one is really in some sense talking about an individual concept. I wish to emphasize again this is not some philosophers' notion invented by Frege and Carnap, but is (according to me) a very standard scientific notion. I think that this is incontestable in probability theory, which wears the formalization right there on its sleeves. It is much less so in mathematics where another formalization is preferred by mathematicians, even though they intuitively use something that would be formalized with individual concepts, at least naturally so, when they teach elementary calculus; or when it is applied to other subjects such as physics, or even to other fields such as economics.

Some of my enthusiasm for the individual concept reading of the calculus dampened when I thought: would I really suggest that students be taught this formalism? The trouble is, I do not think that the conventional formalism with derivatives of functions and so on could really be abandoned in favor of this. I've been calling the conventional view the one which sees derivatives as being over functions, and the individual concepts are really functions too, of course; but they are functions of a very special kind. In fact, this probability textbook says: variables in mathematics are just functions; if you think about a variable quantity, that's what a random variable is. This sort of carries this further, to the case of any scientific quantity. But it would be a lot of work to teach this to freshmen in calculus, or senior high school calculus; and not only that, but you'd have to say, there's this other formalism too, you have to learn it too, and so on. I think similar obstacles face the nonstandard analysis approach being taught to elementary students at the present time; but here it's even more, because I would not propose that we abandon the idea that derivatives are derivatives of real-value functions on the reals, complex-value functions on the complex numbers, and so on. On the contrary, a lot of the mathematics just is that; and then, a mathematician is likely to react that if one learned this new formalism, one would mostly forget it. What the mathematician does heuristically is, often needs it, thinks in terms of this other formalism informally, and is aware of a rough and ready way of translating it into the function formalism without ever having worked out a detailed rule, like translating ordinary language into logic.

Frege pointed out that when you have a description embedded in an intensional operator, it safely can be regarded as about an individual concept. Indeed, if one quantified in in a straightforward way, and wasn't talking about the individual but about something that was not necessarily rigid, such as in the statement, "it is highly probable (probability *much* larger than $1/2$) that the-first-electron-that-hits-the-plate-or-the-last-electron-if-none-do will hit the plate," that is an intensional statement, and all one could say if one made an existential quantification in is, "there is an individual concept $y$ such that it has this probability."

Russell already pointed out that if one wasn't worried about quantifying in but just the original statement itself—actually Frege didn't mention quantifying in either; it was later commentators on him, such as Church and Carnap, who later asserted this—but if you don't worry about quantifying in, then you certainly don't need the quantifiers over individual concepts, and in fact the definite description itself doesn't have to be thought of as denoting an individual concept; one could replace

it (as Russell preferred) by ordinary quantifiers over individuals, buried under the inside of the operator. What we really wish to show is that this trick, acceptable to Russell, will be applicable to the entire Carnap approach. Even where a statement is intuitively about individual concepts, it need not necessarily involve quantification over individual concepts exclusively. This can be done in natural language cases too, in a rough and ready way. What we want to show is that this applies to the entire system.

As long as all quantifiers are on the inside of a modal operator, it is (as Carnap observes) a matter of complete indifference whether they be thought of as over individuals or individual concepts. The results will come out exactly the same way. Moreover, nothing unacceptable to Quine will happen with this kind of quantification, because there is no quantifying in.

So with no intention of getting this particular result, when considering a problem about individual concept quantification, I found that this is what falls out of the theorems. Of course in some sense Kaplan showed—or really, simply observed—that Frege's ontology can be mirrored in Russell's; obviously the sense of such a thing as *the F* could be identified with the unit property of *being F*, which is acceptable to Russell.[16] But here we want to do something more direct, and more like what Russell would accept. Of course, we can't get everything acceptable to Quine, because he finds the modal operators objectionable in themselves, even if every logical hurdle is cleared. I am poking fun, of course, but I am not really trying to get Quine to accept anything. In fact, I myself like the original systems for which these tricks are inapplicable.

Still, we can do a little better. The surprising thing is that individual concept quantifiers can themselves disappear with respect to quantification over individuals plus another type of quantification, or in some cases, with respect to quantification over individuals simpliciter.

Before saying how to do this, I should make a few comments on the two types of individual concept. I have always been talking here about unstructured individual concepts; so that if two individual concepts coincide over all worlds they are the same. But one can also work with structured individual concepts for which this is not the case. I shall not here be considering quantification over structured individual concepts, or structured individual concepts at all, although they may be the relevant entities for some purposes. To take one case where that kind of thing might be relevant: it often would be implausible to regard two complicated mathematical expressions that happen to determine the same number necessarily as identical in meaning. And there might be other applications outside of mathematics as well. Fine's arbitrary objects, as he sets them up, are a certain type of structured individual concept.

But also, many philosophers have probably thought not of quantifying over all individual concepts—here the quantifiers range over all individual concepts, that is, all the lines you can draw across in a given structure.[17] But you might just consider

[16] How to Russell a Frege-Church. ⟨⟨Written in margin; see Kaplan (1975).⟩⟩

[17] Individual concepts are sometimes referred to in the literature as "world-lines" (see, e.g., Thomason, 1969, p. 136), which Kripke may be alluding to here. (The Editors.)

a family of individual concepts, or a family of individual concepts with restrictions, the most natural restriction being, say, that each individual in the domain of each world should be "hit" by at least one individual concept. If it isn't, it really won't appear in the quantification at all.

I really thought about that case way back in the late '50s and announced its solution in an abstract.[18] I don't so much like the way I proved it. (I sort of proved it directly, by modal arguments.) Taking the philosophical point of view I have now, quantification over individual concepts is just a special case of quantification of individuals of a certain sort; and in fact, this case can be reduced to the theorem for the set of individuals, where here "=" appears not as an identity predicate but as a particular equivalence relation between objects being quantified over, such that all the atomic predicates are indiscriminable with respect to this predicate, but not necessarily the ones with modal operators. This observation can be used, with further argument, to reduce the case of quantification over such a family to that of individuals, and derive all completeness and compactness and whatever theorems for the second case from the first. We then can use this case (some of the time, with additional tricks) for the special case where the family is the family of all individual concepts. This is the idea.

I would say that the idea of quantification over a family is the point of view even of some philosophers writing today, who think of it as quantification of individuals with special principles of identification over them across worlds and so on, where even which principles one uses may depend on one's interests and whatever. These are quantifications over families of individuals from the present point of view. But we are now talking about the harder case of quantification over all individuals. Thomason asked whether completeness theorems and axiomatizability results hold for this type of quantification also. In fact, someone I know told me that he had solved this problem negatively; and I accepted this result for many years without thinking about it. I was then unable to reconstruct this for a class, so I called the person up; and his reply was that he himself was not sure that his result was right.

So I looked into this problem. In fact, the answer depends on the underlying modal logic. We have any given underlying normal modal logic—where we define a normal modal logic in the way that is usual today, so that the weakest one is the system **K**—so the semantics is given by an arbitrary relation over a structure $\langle G, K, R\rangle$. The classical Carnap case is the **S5** case. For that case, axiomatizability does indeed hold, but in general the other result proves to be correct.

Here is the basic result. Take the domain to be the set of all individual concepts, and let the atomic predicates be predicates over individuals which induce predicates on individual concepts. If we have **S4**, identity between individual concepts will be expressible as necessary "identity" (i.e. $\Box(x = y)$); if we have **K4**, we can use $(\Box(x = y) \wedge x = y)$. Otherwise we can't do it except by an infinite conjunction of necessity operators. In fact, we might define "$\Box^n A$" inductively as $\Box^0 A =_{\mathrm{df}} A$, and

[18] Presumably, he is referring to Kripke (1959b). This result is not really explicit in that abstract, but it is perhaps obliquely reported in the following: "If identity is added, completeness theorems can be derived either on the assumption $(x, y)(x = y \supset \Box x = y)$ or without this assumption. The resulting semantical notions […] provide a semantical apparatus for sense and denotation, extension and intension, and related concepts" (Kripke, 1959b, p. 324). (The Editors.)

$\Box^{n+1}A =_{\text{df}} (\Box^n A \wedge \Box\Box^n A)$. The first conjunct in this definition is so that even if we don't have $\Box A \supset A$, each $\Box^{n+1}A$ entails $\Box^n A$. So as $n$ rises, '$\Box^n$' is a stronger and stronger operator, but only the conjunction of $\Box^n(x = y)$ for all $n$ will entirely define identity. A model will satisfy this only if it is what they now call a *generated* structure, and I originally called a *connected* structure. With this in place, we can introduce quantification over propositions. We need not worry here about the principle of individuation for these propositions, but we could regard them simply as sets of worlds. At any rate, any set of worlds will be thought of as determining a proposition, so when we quantify over propositions, we quantify over arbitrary sets of worlds.

The general theorem is that if you have any formula $A$ of the Carnap system (as I am now defining it), it can be effectively transformed into a strictly equivalent formula $B$ such that all the individual concept quantifiers are now on the inside of the modal operators, and such that their equivalence is valid on all structures of the appropriate type, satisfying the appropriate axioms. And $B$ can be found effectively from $A$: there is an effective transformation of each formula $A$ into $B$. And $B$ has no quantifying in of individual concepts into the formula at all: all quantifiers are on the inside, so they could just as well be quantifiers over individuals.

However, propositional quantifiers are allowed, and they can quantify in. When I speak of acceptability to Russell, I mean that Russell would have had no objection to such a formula at all, and they wouldn't quantify over any entities he wouldn't immediately recognize as his own. And perhaps even the modern skeptics of modal logic would have regarded quantification over propositions as something likely as part of the rules of the game, and quantification over individual concepts as something more unexpected, which some of them (on the basis of arguments I consider to be fallacious) thought one was forced to use.

This theorem doesn't in itself satisfy the axiomatizability problem because it just reduces it to the case of axiomatizing this. However, given the amount of leisure time that I have in fact been given, perhaps I shall say something as well about how the proof goes. This proof is by induction on the complexity of the formula $A$—as almost all proofs of this type are going to be—and it actually proves something a bit stronger: the formula $B$, depending on the number of nested modal operators in $A$, can be taken to be $n$-monotonic for some $n$, where (if I'm remembering this correctly) $n$ is simply the number of nestings of modal operators that occur in $A$.

As I say, $B$ involves no quantifying in. That really means that there is another formula related to $B$ involving $m$ propositional free variables (in addition to the quantifiers over propositions included in $B$ itself) such that $B$ is a result of replacing each of the formulae $B_1, \ldots, B_m$ by $A_1, \ldots, A_m$, where $A_1, \ldots, A_m$ are formulae of ordinary quantification theory. In other words, these are atomic proposition letters, so they occur way on the inside of all operators and everything, and the assertion is that you can get it in such a way that each of these is replaced by a purely extensional formula. So another way of looking at it is, you have an $m + 1$-tuple of formulae, one of the system of just propositional logic and quantification over propositions, and the

others of extensional classical quantification theory. And then this formula $B$ comes by making this substitution here. Moreover, the first of these formulas $B_1, \ldots, B_m$, can be done in such a way as to be $n$-monotonic—that is, there is some $n$ such that the following is valid:[19]

$$(P_1)\ldots(P_m)(P_1')\ldots(P_m')\Box^n(P_1 \subseteq P_1' \wedge \ldots \wedge P_m \subseteq P_m') \supset (B(P_1, \ldots, P_m) \supset B(P_1', \ldots, P_m')) \quad (10.10)$$

We mean by this the conjunction of all these for each $P_i$, $P_i'$ is another variable. So what we mean is, $P_i$ *strictly strictly strictly* implies $P_i'$. Then if $B$ holds of each $P_i$, it also holds of each $P_i'$.

This isn't too important, but it's useful for the proof. The details of this proof come from a more classical theorem by Feferman and Vaught in pure extensional logic in the theory of direct products, and I followed the exposition in Chang and Keisler's model theory textbook on this.[20] They don't express things this way, obviously, but they have something rather similar to this. The proof will go by induction. For an atomic formula $A$, there's nothing to prove because $B$ is already of the proper form: $A$ has no modal operators. Similarly the conjunction of two formulae of the proper form cannot violate the appropriate form. The negation of a formula in proper form can violate this form, if you want to make sure that it remains monotonic; if you don't worry about that it doesn't. But if you apply negation, the formula will be antimonotonic. But then by using some tricks, applying DeMorgan Laws, pushing the negations in, and replacing the $P$s by their complements, we can make them monotonic again. (Monotonicity isn't a property I'm interested in here *per se*; it's just that it is used in the proof.)

The really tricky case is the existential quantifier. This is where quantification into modal contexts enters the picture, so it is where the whole thing happens. So, you have an existential quantifier:

$$(\exists x)A \wedge (\exists x)B \quad (10.11)$$

This is very much not of the proper form.

An old observation about the individual concept system, not noted by Carnap but pointed out by other modal logicians, was that this formula schema

$$\Box(\exists x)A \supset (\exists x)\Box A \quad (10.12)$$

which under the ordinary individual interpretation is obviously false, under the individual concept interpretation (with all individual concepts available in the domain) is true, provided $A$ contains no modal operators or anyway is extensional. This is because the antecedent says that it is necessary that there be an individual concept

[19] In the transcript, (10.10) was $(P_1)\ldots(P_m)m\Box^n(P_1 \subseteq P_1') \supset (B(P_1, \ldots, P_m) \supset B(P_1', \ldots, P_m'))$, which we have corrected following the suggestion of a referee. (The Editors.)

[20] See Chang and Keisler (1990). (The Editors.)

satisfying $A$, that is, in each world you are assured of being able to choose an individual concept $x$ such that $A$ holds of $x$. Since $A$ is extensional, $A$'s holding of $x$ really depends only on the individual involved, not the entire individual concept $x$. So you could just as well say that in each world you can choose an individual which satisfies $A$. Then take the individual chosen for each world, and make that into a grand individual concept of its own: a function mapping each possible world to the individual in that world which satisfies $A$. This new individual concept will satisfy $A$ in all worlds. And so (10.12)'s consequent will be satisfied whenever its antecedent is.

(10.12) is very strange looking; it is also very useful. A hint as to its usefulness is its apparent similarity to the axiom of choice. There is one step of the argument in particular for which we will need it. Without going through all of the details of the proof, it will be impossible to say exactly how this goes; and of course, in giving the exact details, one would have to take another hour or so, and even if I had the time, it might be a little too complicated to give in this lecture format. (Actually, we have a formula with this variable being the one which is existentially quantified into. The whole thing has been put into a form where there are various things substituted for the propositional variables like this which have no modal operators themselves or are all inside all modal operators. On each world, a certain number of these $A$-formulas may hold. On that world, something else holds that is the existential quantification of the conjunction of all of these. Only some of the formulas may hold, but of those that do hold, the conjunction of their existential quantification holds. Also you can go backwards: if someone is interested in this, maybe it could be brought up in the question period at the end.)

This can be given by propositions in such a way as to set up, that on these privileged worlds, wherever there exists an $x$ satisfying some of these formulae separately, there is an $x$ satisfying their conjunction. Not on all worlds, but on a certain privileged set of worlds: here the privileged set of worlds are those on which the unquantified formula $Ax_1 \ldots x_n$ holds, because if any of these holds on it then their conjunction holds trivially, and that implies the existential quantification of their conjunction. Given that fact, and expressing it in the propositional language, one gets an equivalent formula to the original one with only the propositional quantifiers and the quantifying in.

We also have the following theorem that results: for any modal propositional logic $\mathbf{L}$, let $\mathbf{L}^p$ be the system obtained by adding quantifiers to $\mathbf{L}$. (We can consider the set of valid formulae of $\mathbf{L}$.) Using our main result, one can prove that if $d_{ic}$ is the degree of unsolvability of the set of valid formulae with the system of individual concepts here, the Carnap system, $d_{ic} = d_p \cup 0'$.[21] What this means is, in other words, if you had an oracle that could tell you whether a given formula with quantifying over propositions only and nothing else, a formula of pure propositional calculus augmented by these quantifiers, and had an oracle for quantification theory, you would get an oracle for

[21] In the transcript, the equation appears to be $d_{ic} = d_t \cup 0'$, which we have corrected following the suggestion of a referee. (The Editors.)

quantification over individual concepts out of these two. Given these two decision procedures you can get this decision procedure.

It's actually a little more than that; in fact, it's quite a bit more than that. First, the proof actually implies that the set of all valid formulae here is recursively enumerable. This set in ordinary quantification theory is, of course; that's what Gödel proved. (He was anticipated by some other people like Löwenheim, but Gödel finally proved it in a way that people finally officially accepted, and knew what they were doing.) Suppose that the set of valid formulae of quantified propositional logic is also recursively enumerable. The result would imply that the set of valid formulae for individual concepts is also recursively enumerable—which is not stated here, which is only about degrees of unsolvability, so it is a bit too weak. And vice-versa, by the way, also holds: the interesting thing here is the reduction of quantification over propositions to quantification over individual concepts. This is actually easy. You can come very close to expressing quantification over propositions in the system, but not quite. Suppose the following is the case:[22]

$$\bigwedge_{0 \leq m \leq n} \Box^m (\exists x)(\exists y) x \neq y \tag{10.13}$$

We can express this as "all possible worlds have at least two individuals." In other words, there is no world with only one individual. Then we can as usual express a proposition by an indicator: given an individual concept, consider the concept which coincides with it when the proposition is supposed to be true, and differs from it when the proposition is supposed to be false. Any proposition can be so represented in terms of a given individual concept. In this way we can easily quantify over all propositions, provided that there are at least two individuals in each world. However, we can't quite do this in some cases, where there may be some world with only one individual, which I have not excluded. Nevertheless, so far as the reduction is concerned with validity, then a formula—well, let's take the case where **S4** holds, so we don't have to worry about this necessarily-to-the-$m$ stuff. Then a formula of the form

$$\Box(\exists x)(\exists y)(x \neq y) \supset B^* \tag{10.14}$$

where $B$ is a formula with quantification over propositions, and even individuals and individual concepts too; then we transform it into a formula where the propositional quantifiers are replaced by quantifiers over individual concepts using the obvious indicators here. We can just introduce another free variable, or else say that for all $x$ something is true. Then this will be valid iff the formula with quantification over propositions—or even propositions and individual concepts, of the language you might call $\mathcal{L}_{p,ic}$—is valid. I might say that this language also has exactly the same axiomatizability problem as this and therefore this. But the reduction doesn't quite

[22] In the transcript, (10.13) was $\bigwedge_n \Box^m (\exists x)(\exists y) x \neq y$, which we have corrected following the suggestion of a referee. The use of both $m$ and $n$ is required for the general case (note the simplification alluded to later where **S4** obtains). (The Editors.)

give a translation, because of the necessity of excluding models with worlds with one thing only, which has always been a tricky and delicate point in some of my arguments. Also an upsetting one, but I won't have to worry about that one again.

For validity anyway, the reduction is obvious, and for satisfiability; the trickier part is the type of reduction the other way. In fact it's true—I think; I tried to reverify this over the weekend—that the system with quantification over individual concepts, and therefore even the system with quantification over propositions and individual concepts that reduces to it, is axiomatizable iff the pure propositional system and the purely extensional system are both axiomatizable. The second of these problems was solved by Gödel, as I said, so we can leave it out. So it reduces to the question of axiomatizability of the system with pure quantification over propositions. I do not know, however, any case where this stronger statement is relevant. Every case that I know of either—I mean, one case would be, just consider standard normal finite axiomatizable modal logics **L**. Are there any of them where the pure propositional calculus is recursively enumerable but not recursive? I do not know. In fact, in every case I know of in modal logic where it is relevant, either $d_p$ is equal to 0 (that's what's true in **S5**, which was nice, so the system is axiomatizable) or $d_p$ is equal to—let me just call it "**S0**"—the degree of valid formulae in full second-order logic. And it's really not just a question of my knowledge or lack thereof; this easily transforms into the problem of monadic second-order theories in ordinary logic; and that's another reduction. Though the recursive ⟨⟨~~functions~~ function theorists⟩⟩ here tell us that there is no highest degree of unsolvability and no hardest problem, I think in effect this is the highest degree of unsolvability: all mathematical problems, such as whether the continuum hypothesis is true and so on, that one naturally formulates, reduce to problems about membership in this set. So this is as bad as something could be, in practice.

So in these cases, the result will show that you have the same degree of unsolvability as second-order logic, and probably even could be shown to be recursively isomorphic to second-order logic; that's a stronger result. But I'm saying this off the top of my head, so I may be wrong. Many of these results should really be better formulated, to recursive isomorphism results. However, I will say someone who had heard of my work, and had proved that something else was equivalent to second-order logic—well, actually, the referee of the paper proved that—said, it should be an interesting open problem whether Kripke's results can be extended in the case where they're unsolvable to showing that the systems have the full power of second order logic. But that's just because I wasn't asked, because I had already proved that. I should also mention that Hans Kamp did some kind of reconstruction of my results when they were circulated around in letter form.[23]

For any system **L**, as long as **L** is contained in **S4** or **B**, or in other words, as long as the relation involved is either transitive or symmetric, one has the degree of unsolvability of full second order logic here. So the old result was right; though not for the right reasons here. So here the situation is as bad as it could be. Let me write this equation:

[23] See Kamp (1977). (The Editors.)

$$d_{p,ic} = d_p \cup 0' \tag{10.15}$$

A case where $d_p = 0$ is that of quantified **S5**. Quantified **S5** over propositions is really equivalent to, as a system, ordinary second-order monadic extensional logic, with the first-order operators being the worlds themselves. And in fact, the reductions that can be carried out are probably easier directly applied to the modal language, than to the original classical language that it's imitating; but anyone who knows the first result is likely to get the second, the modal result. At any rate the modal result here, on quantifying over propositions in **S5**, should be attributed to Kit Fine. This really, as he wouldn't deny, is the classical result simply translated, or equivalent to the classical result.

It immediately follows, then, that the system is axiomatizable—because I don't mean just that it has the degree $0'$, that's why the thing is actually recursively enumerable in this case. The reductions are positive; probably recursive isomorphism or something like that holds. At any rate, that's the result.

To show that something is axiomatizable is not to give an elegant axiomatization of it; that shows it is recursively enumerable. But for example in the case of the quantifier there are uncountably many $x$s; first the axiomatizability was shown, and then it was an open problem to give a naturally elegant axiomatization. I mean, axiomatizations can be given in terms of this; you all know Craig's Theorem in the philosophy of science,[24] where he shows how to give you an axiomatization, with these axioms such as "$A \wedge A \wedge A \wedge \ldots$" $n$ times, and so on; but one doesn't want to use that unless one is really stuck. One might like natural proof procedures and so on. At any rate it is a separate problem requiring separate arguments to give an axiomatization of this system for **S5**.

I don't know any modal propositional logic where you don't get either the best or the worst possible result, that it is either $\langle\langle$~~recursive~~ recursively enumerable$\rangle\rangle$ or full second-order logic. (This may be ignorance on my part; some expert on second-order logic might be able to produce some.) One case where I think it is probably still not known whether it is a counterexample or not (although I am not even sure of that!) is the logic **S4.3**.[25] (Saharon Shelah had a result using the continuum hypothesis, or alternatively, Martin's axiom, but anyway using independent things in set theory, that this had a degree greater than or equal to all the truths of arithmetic. But in **ZF** set theory at that time, and as far as I know still now, the status [[of these independent things]] was unknown. And that will be the same thing as the axiomatizability for these things and these, for the ordinary modal system **S4.3**, will depend on these things in set theory.)

But back to **S5**: the following axioms are sufficient. You take the ordinary modal system, as conventionally given in my original paper on quantified **S5**, formulated in such a way that the Barcan Formula and its converse are heaped into the formalization—something I otherwise don't like to do, but which should be done

[24] See Craig (1953, 1956). (The Editors.)

[25] Garson (1984, Sect. 3, pp. 289, 301) claims that **S4.3**-**Q2** is unaxiomatizable, a result he attributes to Kripke. (The Editors.)

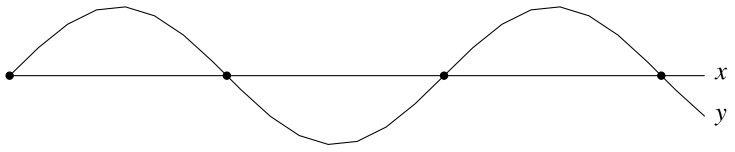
here. This doesn't mean that we really have a constant domain in this case; it just means that the quantifiers are over all individual concepts, which is itself constant. Then the axioms are very natural. These are the same as the ones I gave in the abstract for this talk.

A1 $(x)(x = x)$
A2 $(x)(y)(x = y \supset (Fx \supset Fy))$, where $F$ is free of modal operators.
A3 $\Box(\exists x)A \supset (\exists x)\Box A$, for modal-free $A$.
A4 $(x)(\exists y)(x = y \wedge (z)(x = z \supset \Box(x = y \supset x = z)))$

The first, of course, is straightforward. The second is the substitutivity of identity (that is, of coincidents) so long as modal operators are not present. The third is a sort of funny axiom of choice; really we need it as an equivalence, but the right to left direction is easy to derive, so you need postulate it only in one direction.

Then the fourth thing is a funny doohickey; it's hard even to read what it means. But it turns out that it gives us what we need. (Remember that in this individual concept interpretation, "$x = y$" means that $x$ *coincides* with $y$.) The idea is that $y$ is supposed to be an individual concept which does actually coincide with $x$, but otherwise it tries to avoid $x$ as much as possible. If it is possible to avoid $x$, it does so; so it coincides with $x$ only when there's only one individual. Now it may not even be clear that this axiom says that. And in fact, if the quantifiers range over only a family of individual concepts, it may not say that—something I sometimes in my thinking have to remind myself of. But if the domain includes all possible individual concepts, $y$ will satisfy the second conjunct of (A4) only if it does avoid $x$ so far as possible. That is, any other individual concept which also coincides with $x$ will be such that if $x = y$ then $x = z$. This is because $x = y$ is possible only if there's only one individual; so then $x$ had better equal $z$ too, because there's only one individual in the given world. So that's why this is true. You have to look at it to see that it sets this condition on $y$, but if you parse it out and think about it, you'll see that it does.

This can be illustrated as follows:




The straight line here is the individual concept $x$; the individual concept $y$ avoids it, except in some places where there's only one individual, where it has to hit it. Other than the leftmost dot, which is supposed to be the actual situation, the dots represent worlds with only one entity in the domain. If there are two individuals in each world, then really you can give the statement "$x = y$" as the indicator of the actual world. I've been speaking of the actual world, but since we have a rule of necessitation, (A4) will be necessary, so it will be true with any given world taken as the focus world, so to speak.

The proof that this set of axioms is complete goes through a couple of stages. It might be done more directly inside the modal language; but the way I like to do it is first to consider a family of individual concepts satisfying these axioms. Then I show that the argument I gave before for the reduction of propositional quantifiers and no quantifying in by individuals can be carried out just by the assumption that the family of concepts satisfies these axioms. I then show that if you have a family of concepts satisfying these axioms, it will be elementarily equivalent in the modal sense to the full family. Using these facts—and that really involves—what properties of the class of all propositions do you need here? It turns out that all you need really is that every singleton world can be picked out. That isn't quite guaranteed by this, but in a certain normalized class of relevant structures, it will be guaranteed.

For this case, given known reductions, we can in fact in **S5** get rid of the propositional quantifiers altogether and replace the system by one in which you have only operators $\Diamond^m A$—meaning "for at least $m$ worlds, $A$ holds"; ordinary possibility is $\Diamond^1$—otherwise no quantifying in occurs. The only quantifiers then can be thought of as over individuals, and they are on the inside. So if there are free variables (as you should remember with the interpretation of this thing) they still will have an intensional character to a degree, because "$x = y$" and "$\Box x = y$" won't necessarily mean the same thing here. But if there are no free variables, you might as well think of the whole system as a system about individuals.

You're going to have to have axioms for this system, and I don't at the moment know the most elegant way to do it. I proposed one way in the JSL. As I stated there,[26] but it seems to me I was looking at it in too complicated a way, if you're willing to strengthen the system, one can get rid of the $\Diamond^m$ operators also in favor of an ordinary possibility operator. The idea of this is simply that you can consider only structures in which there are infinitely many isomorphic copies of every world. This does not affect the logic if this is the logic of ordinary quantification over individuals with modal operators, because duplicate worlds might as well be identified. On the other hand, in this transformation, $\Diamond^m$ will therefore be equivalent to $\Diamond^1$ for any closed formulae. In the thing in the JSL, I also wanted to have this be the case for open formulae, so I also say there has to be a restriction on the family of individuals involved. But that adds a complication; let's say we're concerned only with interpreting formulae without free variables.

We already have in the general case that for closed formulae quantifiers over individual concepts can be replaced by quantifiers over individuals that never quantify in, and quantifiers over propositions. In the **S5** case we can eliminate the quantifiers over propositions in favor of these $\Diamond^m$ operators, so any closed formula can be interpreted (even in the original system with quantification over individual concepts) as if it had only these operators and quantifiers over individuals. With the strengthened model-theoretic assumption that every world repeats infinitely often, we get some-

[26] [[Actually it is not obvious from Kripke's actual words that one and the same JSL publication contains the axiomatization and the method of eliminating $\Box^m$ $\langle\langle(\Diamond^m)\rangle\rangle$ operators.]] Presumably this is the same as the paper from the 1950s discussed on p. 27 (Craig's Theorem) and earlier, "I sort of proved it directly", CHECK! $\langle\langle$(written in the margin)$\rangle\rangle$.

thing which is (as far as closed formulae are concerned) a conservative extension of the original system with quantifying over individuals—or with quantifying over individual concepts, because it makes no difference, since all the quantifiers are on the inside. Yet every closed formula, every sentence, is equivalent to a sentence in this system. That is, the statement that says they are equivalent is valid.

That means in a certain sort of way we could look at this in a different way; and if you want to allow propositional operators you could look at it this way for the general case. We could say that quantification over individual concepts is only a *façon de parler*, and we don't have to invoke these entities at all given these restrictions. It is replaced by quantification over propositions and antihaecceitist quantification over individuals; or even, in this case, without quantification over propositions, and with this operator, $\langle\langle\Diamond^m\rangle\rangle$; or in the even stronger case, with only the ordinary possibility operator.

In this way the Carnap system (e.g.) can be interpreted in the system with only quantification over individuals; but really the neutral system, since there is no quantifying in, the issues which distinguish individuals from individual concepts are not really dealt with—the question of the necessity of identity or anything like that. One can interpret the full Carnap system inside this system that has no quantifying in. From a certain philosophical point of view that might be a very satisfying result; and aside from one's philosophical point of view, it is a somewhat nice-looking and surprising result. And we get stronger and stronger results in this direction.

It is interesting to investigate whether there are other cases where one gets a good result. Tense logic for Prior's model of discrete time, I think, will be one where one gets a good result, somewhat similar to **S5**—at least, the valid formulae are recursively enumerable. But it's very likely that in a weaker system than **S5** one gets a bad result. I should have said that these results also extend if you have tense operators or things; you know, for richer modal operators, in some cases the same techniques will work, at least for many modal operators.

Perhaps I should stop right here. There are other things, both about the application of individual concepts and about the theorems which I haven't dealt with, but this is obviously more than enough. So I hope you see why, in spite of my advance philosophical prejudices against the birth of these systems, I really very much like some of the things about the way they came out, both in the way ideas based on them can be used, and in theorems about them.

## References

Bacon, J. (1980). Substance and first-order quantification over individual-concepts. *Journal of Symbolic Logic, 45*(2), 193–203.

Barcan, R. C. (1946). A functional calculus of first order based on strict implication. *Journal of Symbolic Logic, 11*(1), 1–16.

Carnap, R. (1958). *Meaning and necessity: A study in semantics and modal logic*. Chicago: University of Chicago Press.

Chang, C. C., & Keisler, H. J. (1990). *Model theory* (3rd ed.). Amsterdam: North-Holland.

Craig, W. (1953). On axiomatizability within a system. *Journal of Symbolic Logic, 18*(1), 30–32.
Craig, W. (1956). Replacement of auxiliary expressions. *Philosophical Review, 65*(1), 38–55.
Evans, G. (1985). Pronouns, quantifiers, and relative clauses (I). In G. Evans (Ed.), *Collected papers* (pp. 76–152). Oxford: Clarendon Press.
Feller, W. (1968). *An introduction to probability theory and its applications* (3rd ed., Vol. 1). New York: Wiley.
Fine, K. (1970). Propositional quantifiers in modal logic. *Theoria, 36*(3), 336–346.
Fine, K. (1985). *Reasoning with arbitrary objects*. Oxford: Blackwell.
Garson, J. W. (1984). Quantification in modal logic. In D. Gabbay & F. Guenthner (Eds.), *Extensions of classical logic* (1st ed., Vol. 2, pp. 249–307). Dordrecht: D. Reidel Publishing Company.
Hughes, G. E., & Cresswell, M. J. (1996). *A new introduction to modal logic*. London: Routledge.
Kamp, H. (1977). *Two related theorems by D. Scott and S. Kripke.* (Mimeographed.)
Kaplan, D. (1975). How to Russell a Frege-Church. *Journal of Philosophy, 72*(19), 716–729.
Kripke, S. A. (1959a). A completeness theorem in modal logic. *Journal of Symbolic Logic, 24*(1), 1–14.
Kripke, S. A. (1959b). Semantical analysis of modal logic (abstract). *Journal of Symbolic Logic, 24*(4), 323–324.
Kripke, S. A. (1985). Review of Kit Fine, 'Model theory for modal logic. Parts I–III' (Journal of Philosophical Logic, 7 (1978), pp. 125–156, 277–306; and 10 (1981), pp. 293–307). *Journal of Symbolic Logic*, *50*(4), 1083–1093.
Kripke, S. A. (1992). Individual concepts: Their logic, philosophy, and some of their uses (abstract). *Proceedings and Addresses of the American Philosophical Association, 66*(2), 70–73.
Quine, W. (1976). Whither physical objects? In R. S. Cohen, P. K. Feyerabend, & M. W. Wartofsky (Eds.), *Essays in memory of Imre Lakatos* (Vol. 39, pp. 497–504). Dordrecht: Springer.
Thomason, R. H. (1969). Modal logic and metaphysics. In K. Lambert (Ed.), *The logical way of doing things* (pp. 119–146). New Haven: Yale University Press.
Thomason, R. H. (1970). Some completeness results for modal predicate calculi. In K. Lambert (Ed.), *Philosophical problems in logic: Some recent developments* (pp. 56–76). Dordrecht: D. Reidel Publishing Company.

# Chapter 11
# Nested Sequents or Tree-Hypersequents—A Survey

**Björn Lellmann and Francesca Poggiolesi**

**Abstract** This paper presents an overview of the methods of nested sequents or tree-hypersequents that were originally introduced to provide a comprehensive proof theory for modal logic. The paper retraces the history of how these methods have developed. Its aim is also to present, in an unified and harmonious way, the most recent results that have been obtained in this framework. These results encompass several technical achievements, such as the interpolation theorem and the construction of countermodels. Special emphasis is also given to the application to logics other than the standard modal ones as well as to relations to other proof theoretic formalisms.

**Keywords** Cut-elimination · Decidability · Interpolation · Modal logic · Proof theory

## 11.1 Introduction

The sequent calculus, introduced in Gentzen (1935), was designed to construct derivations for classical (and intuitionistic) logic and to prove that any of these derivations can be made *analytic*, where a derivation is analytic when each formula that occurs in it is a subformula of its conclusion. Several successive studies have shown that the classical sequent calculus **Gcl** (see Fig. 11.1) displays different properties that have both technical and philosophical value. For example, the logical rules of the classical sequent calculus are[1]:

[1] See Avron (1996), Poggiolesi (2008c, 2010b) and Wansing (2002) for a discussion these properties, as well as for others.

B. Lellmann
TU Wien, Vienna, Austria
e-mail: lellmann@logic.at; bjoern.lellmann@bmf.gv.at

Federal Ministry of Finance Austria, Vienna, Austria

F. Poggiolesi (✉)
IHPST, Paris, France
e-mail: poggiolesi@gmail.com

© The Author(s), under exclusive license to Springer Nature Switzerland AG 2024

Y. Weiss and R. Birman (eds.), *Saul Kripke on Modal Logic*, Outstanding Contributions to Logic 30, https://doi.org/10.1007/978-3-031-57635-5_11

**Axioms:**

$$\frac{}{p \Rightarrow p}$$

**Structural rules:**

$$\frac{M \Rightarrow N}{\alpha, M \Rightarrow N}\ WL \qquad \frac{M \Rightarrow N}{M \Rightarrow N, \alpha}\ WR \qquad \frac{\alpha, \alpha, M \Rightarrow N}{\alpha, M \Rightarrow N}\ CL \qquad \frac{M \Rightarrow N, \alpha, \alpha}{M \Rightarrow N, \alpha}\ CR$$

**Logical rules:**

$$\frac{M \Rightarrow N, \alpha}{\neg\alpha, M \Rightarrow N}\ \neg L \qquad \frac{\alpha, M \Rightarrow N}{M \Rightarrow N, \neg\alpha}\ \neg R \qquad \frac{\alpha_i, M \Rightarrow N}{\alpha_0 \wedge \alpha_1, M \Rightarrow N}\ \wedge L \qquad \frac{M \Rightarrow N, \alpha \quad M \Rightarrow N, \beta}{M \Rightarrow N, \alpha \wedge \beta}\ \wedge R$$

**Cut rule:**

$$\frac{M \Rightarrow N, \alpha \qquad \alpha, P \Rightarrow Q}{M, P \Rightarrow N, Q}\ Cut_\alpha$$

**Fig. 11.1** The classical sequent calculus **Gcl**

- explicit, i.e., the introduction rules for a constant Ψare explicit if they exhibit Ψin their lower sequent only and they exhibit only one occurrence of Ψ
- separated, i.e., the introduction rules for a constant Ψare separated if they do not exhibit any other connective other than Ψ

Moreover, it can be proved that the classical structural rules are admissible, i.e., a rule is admissible when whatever can be proved with the aid of that rule, can also be proved without; see for example Troelestra and Schwichtenberg (1996). Finally, the classical sequent calculus introduced by Gentzen does not contain any explicit semantic element, where we say that a sequent $S$ does not contain a semantic element if every element that serves to define the sequent itself can be translated in such a way that it forms, together with the translation of the other elements of $S$, a formula equivalent to $S$; see Poggiolesi (2010b), Stouppa (2007) and Stewart and Stouppa (2004).

Since 1935, an impressive number of different logics have been developed. Amongst them, one family which has proved to be particularly successful, is that of modal logic, that is obtained by adding to classical logic the two operators $\Box$ and $\Diamond$ for necessity and possibility, respectively. The blossoming of modal logic is directly linked to the introduction of Kripke possible worlds semantics in the middle of the 1960s, based on work done as early as 1959/1960 (see Kripke 1963a, b; 1965); by means of such semantics, the modal connectives $\Box$ and $\Diamond$ are interpreted very naturally. In this interpretation, a formula $\phi$ is necessarily true if, and only if, it is true in every possible word (of a Kripke model), while a formula $\phi$ is possibly true if, and only if, there exists some world (of a Kripke model) where it is true. Since its introduction, Kripke semantics have been modified, extended and analyzed in a plethora of ways and it has given rise to a great number of applications.

These outstanding results soon led to the investigation of the links between the sequent calculus and modal logic. From the 1950s until the 1980s, the first approaches

consisted in adding to the standard sequent calculus **Gcl** for classical logic (see Fig. 11.1) new rules for introducing on the left or on the right of the sequent the modal operators $\Box$ and $\Diamond$, see, e.g., Curry (1952), Goble (1974), Leivant (1981), Ohnishi and Matsumoto (1957) and Sambin and Valentini (1982). However, it did not take long to realize that these new modal sequent calculi no longer enjoyed most of the properties satisfied by **Gcl**: in certain cases, it even turned out to be difficult to have a cut-free sequent calculus. Hence, the problem did not seem to be that of choosing suitable rules for modal logic, but that of finding rules, and calculi, that did satisfy certain properties. In other words, it seemed that the move from classical logic to modal logic could not be simply made by adding a new connective with rules governing it, but by *extending one's conception of the objects to be manipulated by such rules*. Generalizing the sequent calculus in order to obtain an adequate proof theory for modal logic was thus seen as a necessary step.

But how does one generalize the sequent calculus? Consider the main object of the sequent calculus, namely a *sequent*:

$$M \Rightarrow N$$

where $M$ and $N$ are multisets of classical formulae, the *antecedent* and *consequent*, respectively, separated by the *sequent arrow*. Now try to find a more abstract version of it. Then, several possibilities arise[2]:

- one might want to deal with more than one antecedent or one consequent;
- one might want to deal with more than just one type of sequent arrow;
- one might want to deal with more than just one way of separating one formula from another (i.e., beside the comma);
- one might want to deal with more than one sequent at a time.

Each of these possibilities has given rise to a different rigorous way for improving the standard sequent calculus; in particular, they have led to higher-arity sequent calculi (e.g., see Blamey & Humberstone, 1991), multiple sequent calculi (e.g., Indrezejczak, 1997), display calculi (e.g., Wansing, 2002), and the framework which we consider here, introduced independently under the names of *nested sequents* in Bull (1992) and Kashima (1994), *deep sequents* in Brünnler (2006), and *tree-hypersequents* in Poggiolesi (2008b). Arguably, tree-hypersequents or nested sequents are one of the most successful approaches to the proof theory of modal logic: not only because of the several results and applications that have been obtained with them and which will be described in detail in the next sections (e.g., admissibility of the structural rules, cut-elimination, decidability, interpolation), but also because tree-hypersequents or nested sequents seem to best preserve the original spirit of Gentzen sequent calculus for classical logic: for example, these calculi do not contain any explicit semantic element, and the rules are (often) explicit and separated.

[2] Note that here we do not mention all extensions of the sequent calculus, the list being too long. Moreover, we completely omit the sequent calculi for modal logic obtained not by changing the standard sequent, but by indexing formulae of the language of modal logic, e.g., see Kanger (1957), Simpson (1994) and Viganò (2000).

The aim of this paper is to present an overview of the tree-hypersequents or nested-sequents calculi in the proof theory for modal logic and beyond.[3] As far as we know, no such overview exists so far, and given the wide variety of results obtained, it seems to be needed. We have tried to be reasonably comprehensive but due to space constraints we needed to concentrate on some aspects. Of course we might also have unintentionally overlooked some results. The reader is referred in addition, e.g., to the list of publications on nested sequents on the website http://alessio.guglielmi.name/res/cos/index.html.

The paper is organized as follows. In Sect. 11.2 we provide the basic definitions of nested sequent calculi for the standard normal modal logics, we prove the admissibility of structural rules, soundness and completeness plus cut-elimination. In Sect. 11.3 we show how nested sequents can be applied for countermodel construction (Sect. 11.3.1) as well as for proving decidability and interpolation results (Sects. 11.3.2 and 11.3.3). Section 11.4 contains an overview of other logics which have been considered in the nested sequent framework, including intuitionistic and first-order logic (Sect. 11.4.1) as well as other kinds of modal logics such as intuitionistic and constructive modal logic (Sect. 11.4.2), epistemic modal logics and propositional dynamic logic (Sect. 11.4.3), and non-normal modal logics and conditional logics (Sect. 11.4.4). Finally, in Sect. 11.5 we situate the nested sequent framework in relation to other proof theoretic frameworks including standard sequents (Sect. 11.5.1), hypersequents (Sect. 11.5.2), prefixed tableaux (Sect. 11.5.4), labelled sequents (Sect. 11.5.5) and display calculi (Sect. 11.5.7). We also consider the restriction to linear nested sequents (Sect. 11.5.3) and the extension to indexed nested sequents (Sect. 11.5.6).

## 11.2 The Formalism

We use this section to introduce the nested sequents method.[4] We will do it in the following way: we will first recall some basic notions of modal logic which, although quite well-known, will be important for what follows. Secondly, we will explain the basic ideas behind nested sequents; finally, we will introduce the method formally.

**Definition 11.2.1** Let $\mathcal{L}$ be the *modal propositional language* composed of propositional atoms $p, q, r, \ldots$, the logical symbols $\neg$, $\wedge$, $\Box$, and parentheses (, ). The connectives $\vee$, $\rightarrow$ and $\leftrightarrow$ are defined as usual and might be occasionally used as abbreviations. The same hols for the diamond $\Diamond$, which is defined as $\neg\Box\neg$. *Formulae* are build up in the usual way and are denoted by greek letters $\phi, \epsilon, \ldots$. We use the capital letters $M, N \ldots$ to denote multisets of formulae.

[3] We will present only sketches of proofs and give references to the original papers.

[4] From now till the end of the paper, we adopt the name *nested sequent* instead of *tree-hypersequent*, or *deep sequent* since it is historically the first and also the most used one.

**Definition 11.2.2** A *normal modal system* is a set of well formed formulae of $\mathcal{L}$ such that it contains all the classical tautologies and the distribution axiom:

$$\Box(\phi \to \epsilon) \to (\Box\phi \to \Box\epsilon)$$

It also contains modus ponens, uniform substitution and the *necessitation rule* namely the rule:

$$\text{if} \quad \vdash \phi \quad \text{then} \quad \vdash \Box\phi$$

The modal Hilbert system K is the weakest normal modal system: it contains only the distribution axiom and necessitation rule from Definition 11.2.2 above. Other prominent normal modal systems are obtained by adding to K combinations of the axioms listed in Fig. 11.2: they are named according to the concatenation of the names of their axioms. An exception is provided by systems S4 and S5: the former is obtained by adding to the system K the axioms $T$ and 4, the latter the axioms $T$, 4 and $B$. These Hilbert normal modal systems are taken to be the main or standard systems of modal logic and they form the so-called *modal logic cube*, see Fig. 11.2.

Let us introduce some basic semantic notions.

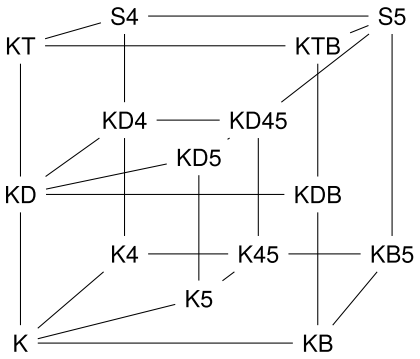



**Fig. 11.2** The modal logic cube together with the axioms and the corresponding frame properties

**Definition 11.2.3** A *frame* $\mathcal{F}$ is a pair $(W, R)$ such that:

– $W$ is a non empty set (of possible worlds), and
– $R \subseteq W \times W$ is a binary relation (the accessibility relation) on $W$.

**Definition 11.2.4** A *tree-frame* $\mathcal{T}$, or simpler a *tree*, is a frame $(W, R)$ that forms a directed, connected and acyclic graph; the nodes of the graph are labelled by the variables $i, j, \ldots$ of the set $W$, and the connection between nodes is established by the relation $R$.

**Definition 11.2.5** A *model* $\mathfrak{M}$ is a pair $(\mathcal{F}, v)$, where $\mathcal{F}$ is a frame and $v$ is the following valuation on $\mathcal{F}$:

$$v := W \otimes PL \rightarrow \{0, 1\}$$

We say that a model $\mathfrak{M} = (\mathcal{F}, v)$ is *based on* the frame $\mathcal{F}$.

**Definition 11.2.6** Given a model $\mathfrak{M} = ((W, R), v)$, $i \in W$, and a formula $\phi$, the relation

$$i \Vdash_{\mathfrak{M}} \phi \qquad [\text{or } v(i, \phi) = 1, \text{ or } \phi \text{ is true at the world } i \text{ of the model } \mathfrak{M}]$$

usually called the *satisfiability relation,* is inductively defined in the following way:

– $i \Vdash_{\mathfrak{M}} p$ iff $v(i, p) = 1$,
– $i \Vdash_{\mathfrak{M}} \neg\epsilon$ iff $i \nVdash_{\mathfrak{M}} \epsilon$,
– $i \Vdash_{\mathfrak{M}} \epsilon \wedge \delta$ iff $i \Vdash_{\mathfrak{M}} \epsilon$ and $i \Vdash_{\mathfrak{M}} \delta$,
– $i \Vdash_{\mathfrak{M}} \Box\epsilon$ iff $(\forall j \in W)\,(i\,R j \rightarrow j \Vdash_{\mathfrak{M}} \epsilon)$.

**Definition 11.2.7** We say that a formula $\phi$ is *true in a model* $\mathfrak{M}$, in symbols: $\Vdash_{\mathfrak{M}} \phi$, if it is true at every world in the model. We say that a formula $\phi$ is *valid in a frame* $\mathcal{F}$, in symbols: $\Vdash_{\mathcal{F}} \phi$, if $\phi$ is true in every model based on that frame. We say that a formula $\phi$ is *valid in a class* $\mathfrak{C}$ *of frames*, in symbols: $\Vdash_{\mathfrak{C}} \phi$, if $\phi$ is valid in every frame which belongs to that class.

Each of the Hilbert normal modal systems forming the modal logic cube is sound and complete with respect to a different class of frames enjoying the property(ies) linked to the corresponding axiom(s), e.g., see Poggiolesi (2010b, p. 44).

Let us now introduce the nested sequents method by presenting first of all the basic intuition behind it. We do that by passing through the semantic link between classical and modal logic. One way of looking at this link is the following: both classical and modal logic evaluate (the truth values of) formulae with respect to possible words, but while in modal logic we have a set of different possible worlds, in classical logic there is only one possible world (which is therefore often omitted).

Let us look at a sequent $M \Rightarrow N$ as describing one possible world; we will use the expression *world-sequent* to denote this perspective. Since in classical logic there is

only one world, only one sequent is required. In modal logic, instead, we deal with $n$ possible words; syntactically, we will then need to deal with $n$ sequents at a time. This is the first step toward a generalization of the sequent calculus. However, as clarified by the previous definitions, worlds obey a structure regulated by an accessibility relation. In the case where the accessibility relation is an equivalence relation (this is what happens in S5), i.e., in the case where the $n$ possible worlds can be looked at as just a set of worlds divided by an equivalence relation, syntactically, we can simply work with a set of $n$ sequents, those in a single equivalence class. A set of $n$ sequents is the object standardly called a *hypersequent*.[5] In the cases where, on the other hand, the accessibility relation amongst worlds enjoys some properties (possibly none), then the $n$ world-sequents need to be arranged somehow. Thanks to the property called the *tree-model property* (which tells us that any satisfiable formula is satisfiable in a model based on a tree-frame, see Blackburn et al., 2001), attention in modal logic can be restricted to tree-frames; consequently, world-sequents can be arranged so as to form a tree. Such a structure is what is called a *nested sequent*, *deep sequent*, or *tree-hypersequent* structure. Hence a nested sequent is simply a bunch of sequents arranged as a tree.

Nested sequent calculi were independently introduced by Bull (1992), Kashima (1994), Brünnler (2006) and Poggiolesi (2008b, 2010b). Different notations come with the different authors, where the main difference is that between a single-sided and a two-sided formulation. Here we adopt the two-sided formulation in a notation which seems to have become standard and can be found in Bull (1992), Fitting (2014) and Ramanayake (2018). See Sect. 11.2.1 below for a brief comparison with the single-sided formulation.

**Definition 11.2.8** (*Nested sequent*) A *nested sequent* is a finite object defined recursively as follows:

- if $M \Rightarrow N$ is a sequent, then it is a nested sequent,
- if $M \Rightarrow N$ is a sequent and $G_1, \ldots G_n$ are nested sequents, then $M \Rightarrow N, [G_1], \ldots [G_n]$ is a nested sequent.

In the following we often use $G, H, \ldots$ to stand for nested sequents. Intuitively then, where a nested sequent can be seen as a tree with each node labelled with an ordinary sequent, the *nesting operator* [.] in a nested sequent captures the successor relation in the underlying tree. From a more syntactic perspective, the nesting operator can be seen as a *structural connective* corresponding to the logical connective $\Box$, in the same way that, e.g., the comma on the right hand side of an ordinary sequent can be seen as the structural connective corresponding to disjunction. This gives rise to the following formula interpretation of nested sequents.

[5] See Sect. 11.4.2 (and also Poggiolesi, 2010c) for the links between nested sequents and hypersequents.

$M \Rightarrow N, [P \Rightarrow Q, [S \Rightarrow T], [Z \Rightarrow W]]$

$M \Rightarrow N, [P \Rightarrow Q, [S \Rightarrow T]], [U \Rightarrow V, [Z \Rightarrow W]]$

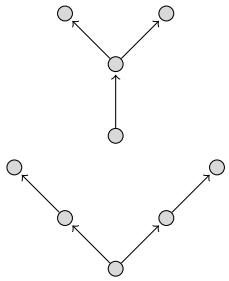


**Fig. 11.3** Two examples of nested sequents with the corresponding tree structure

**Definition 11.2.9** (*Standard interpretation*) The *standard interpretation* $\Delta$ of a nested-sequent is defined recursively as follows:

- $(M \Rightarrow N)^{\Delta} := \bigwedge M \to \bigvee N$,
- $(M \Rightarrow N, [G_1], \ldots, [G_n])^{\Delta} := (\bigwedge M \to \bigvee N) \vee \Box(G_1)^{\Delta} \vee \cdots . \vee \Box(G_n)^{\Delta}$.

Figure 11.3 shows two examples of nested sequents with the corresponding tree-frames of Kripke semantics. Before considering the rules of the nested calculi for the main systems of modal logic, we need to introduce a notational device that allows us to display these rules. Such a device is called a *context* and it is the occurrence of the symbol $\{\}$ in a nested sequent (which does not occur inside formulae). Informally, a context allows us to focus on any part of a nested sequent, where the symbol $\{\}$ acts as a placeholder for a subtree starting at a node of the nested sequent; formally it is defined as follows.

**Definition 11.2.10** The notion of a *nested sequent with a context* is inductively defined by:

- $\{\}$ is a nested sequent with a context of depth 0,
- if $M \Rightarrow N, [H_1], \ldots, [H_n]$ is a nested sequent and $G\{\}$ is a nested sequent with context of depth $m$, then $M \Rightarrow N, [H_1], \ldots, [H_n], [G\{\}]$ is a nested sequent with context of depth $m + 1$.

The result of *filling* a nested sequent with a context $G\{\}$ with another nested sequent $H$ is written as $G\{H\}$ and is obtained by replacing the symbol $\{\}$ in $G\{\}$ with $H$.

Note that the depth of a context in a nested sequent is simply the length of the path from the node labelled with $\{\}$ to the root. Note also that when showing the rules of the calculi, we might abuse notation and only write $N$ instead of $N, [G_1], \ldots, [G_n]$.

As stressed by Poggiolesi (2010b, pp. 8–9) and Troelestra and Schwichtenberg (1996, p. 51), Gentzen systems generally come in *variants*. Consider for example classical logic; there are at least two variants of the sequent calculus for classical logic: the *general* variant and the *logical* variant. The general variant for classical

logic is a sequent calculus where structural and logical rules are taken as primitive and their roles are kept separate. The logical variant (see Dragalin, 1988) is a sequent calculus that is composed by generalized axioms as well the rules $\neg R$ and $\neg L$ and $\wedge R$; as for the rule introducing the conjunction connective on the left side of the sequent, we consider its multiplicative version, namely the rule

$$\frac{\phi, \epsilon, M \Rightarrow N}{\phi \wedge \epsilon, M \Rightarrow N} \wedge L'$$

The logical variant is called in this way because it contains no structural rule: the force of the structural rules is absorbed by the axioms and the logical rules. Variants of Gentzen systems are different sequent calculi; however, since they are different sequent calculi for the same logic, i.e., different sequent calculi that prove the same sequents, it is profitable to think of them as variants of a common core. In what follows, we chose to mainly work with the logical variant of nested calculi. Indeed, not only is the logical variant the closest to the semantic interpretation of modal logic, but also it is through this variant that most of the results can be more elegantly presented.

The nested calculus $\mathbf{NS_k}$ for the basic modal system $\mathsf{K}$ is composed of the axioms and rules given in Fig. 11.4. The rules that introduce negation and conjunction on the left and on the right of the sequent are the standard rules adapted to the nested framework: they state that the classical propositional rules can be applied in any world-sequent. As for the modal rules, they reflect at the syntactic level the satisfaction of a formula of the form $\Box\phi$ in a world $x$ of a Kripke-model. Indeed, the rule $\Box R$ states, if read bottom up, that if $\Box\phi$ is false in a world-sequent $x$, then there exists a world-sequent $y$ accessible from $x$ where $\phi$ is false. The rule $\Box L$, in contrast, states that if $\Box\phi$ is true in a world-sequent $x$, then in any world-sequent $y$ accessible from $x$, $\phi$ is true.

**Initial Nested Sequents**

$$\frac{}{G\{p, M \Rightarrow N, p\}}$$

**Logical Rules**

$$\frac{G\{M \Rightarrow N, \alpha\}}{G\{\neg\alpha, M \Rightarrow N\}} \neg L \qquad \frac{G\{\alpha, M \Rightarrow N\}}{G\{M \Rightarrow N, \neg\alpha\}} \neg R$$

$$\frac{G\{\alpha_0, \alpha_1, M \Rightarrow N\}}{G\{\alpha_0 \wedge \alpha_1, M \Rightarrow N\}} \wedge L \qquad \frac{G\{M \Rightarrow N, \alpha_0\} \quad G\{M \Rightarrow N, \alpha_1\}}{G\{M \Rightarrow N, \alpha_0 \wedge \alpha_1\}} \wedge R$$

$$\frac{G\{\Box\alpha, M \Rightarrow N, [\alpha, P \Rightarrow Q]\}}{G\{\Box\alpha, M \Rightarrow N, [P \Rightarrow Q]\}} \Box L \qquad \frac{G\{M \Rightarrow N, [\Rightarrow \alpha]\}}{G\{M \Rightarrow N, \Box\alpha\}} \Box R$$

**Fig. 11.4** The basic nested sequent calculus $\mathbf{NS_k}$ for modal logic $\mathsf{K}$

$$\frac{G\{\Box\alpha, M \Rightarrow N, [\alpha \Rightarrow]\}}{G\{\Box\alpha, M \Rightarrow N\}}\ d \qquad \frac{G\{\Box\alpha, \alpha, M \Rightarrow N\}}{G\{\Box\alpha, M \Rightarrow N\}}\ t$$

$$\frac{G\{\Box\alpha, M \Rightarrow N, [\Box\alpha, P \Rightarrow Q]\}}{G\{\Box\alpha, M \Rightarrow N, [P \Rightarrow Q]\}}\ 4 \qquad \frac{G\{\alpha, M \Rightarrow N, [\Box\alpha, P \Rightarrow Q]\}}{G\{M \Rightarrow N, [\Box\alpha, P \Rightarrow Q]\}}\ b$$

**Fig. 11.5** The special logical rules for extensions of modal logic K

In order to obtain the calculi for the remaining modal normal systems, we add combinations of the special logical rules given in Fig. 11.5 to the calculus $\mathbf{NS}_\mathsf{K}$. The resulting calculus is written $\mathbf{NS}_\mathsf{K} \cup X$ where $X$ is the set of additional rules. Indeed, each rule corresponds to one of the axioms listed in Fig. 11.2. In the calculi containing the rules 4 and $b$, we also need to add the following rule:

$$\frac{G\{\Box\phi, M \Rightarrow N, [\Box\phi, P \Rightarrow Q]\}}{G\{M \Rightarrow N, [\Box\phi, P \Rightarrow Q]\}}\ 5$$

In contrast to all the other rules, 5 does not reflect the strength and the power of the corresponding axiom: indeed the calculus $\mathbf{NS}_\mathsf{K}$ plus 5 is not cut-free, while the calculus $\mathbf{NS}_\mathsf{K}$ plus the rules 5 and $t$ is not complete. Therefore, the rule 5 just serves to complete the calculi obtained by adding the rules 4 and $b$ to the calculus $\mathbf{NS}_\mathsf{K}$. We call this proviso the *5-closure* of a calculus[6]:

**Definition 11.2.11** A set $X \subseteq \{d, t, b, 4\}$ satisfies the *5-closure proviso* if whenever $\{b, 4\} \subseteq X$, then $5 \in X$.

### *11.2.1 Single-Sided Nested Sequents*

As in the case of standard sequents, nested sequents have also been considered in a single-sided formulation, most prominently in the series of works starting with Kashima (1994) and Brünnler (2006). Similarly to the case of classical propositional logic, for normal modal logics based on classical logics the two formulations can be translated into each other. The benefit of the single-sided formulation is that the presentation of nested sequents and nested sequent rules is much more economic. The downside of this is that formulae are often assumed to be in *negation normal form*, i.e., negation is allowed only immediately in front of propositional variables, and it is because of this feature that we need to assume that we have the dual operators of the modal box, i.e., the modal diamond, as well as of conjunction, i.e., disjunction, in the language.

[6] See Brünnler (2009) and Poggiolesi (2008d).

**Definition 11.2.12** A *single-sided nested sequent* is a structure

$$M, [G_1], \ldots, [G_n]$$

where $M$ is a multiset of formulae in negation normal form, $n \geq 0$ and $G_1, \ldots, G_n$ are single-sided nested sequents. The *standard interpretation* $\Delta$of single-sided nested sequents is given by $(M, [G_1], \ldots, [G_n])^\Delta := \bigvee M \vee \Box G_1^\Delta \vee \cdots \vee \Box G_n^\Delta$.

Thus, in a single-sided nested sequent the nodes of the underlying tree are labelled with single-sided sequents, i.e., multisets of formulae in negation normal form, instead of two-sided sequents. The notions of *context*, *depth*, etc., are defined analogously to the two-sided formulation. The modal rules of the single-sided nested sequent system for modal logic $\mathsf{K}$ then take the following form:

$$\frac{G\{M, [\phi]\}}{G\{M, \Box\phi\}}\ \Box_R \qquad \frac{G\{M, \Diamond\phi, [H, \phi]\}}{G\{M, \Diamond\phi, [H]\}}\ \Box_L$$

The rules for extensions are formulated analogously. We refer the reader to, e.g., Kashima (1994), Brünnler (2006) and Marin and Straßburger (2014) for the detailed description of the systems.

### *11.2.2 Admissibility and Invertibility Results*

In this section we show that all structural rules are admissible in the nested calculi for the main normal systems of modal logic. We will also show that the logical rules are invertible.

**Lemma 11.2.13** (Generalized initial sequents) *For any $X \subseteq \{d, t, b, 4\}$ under the 5-closure proviso, the nested sequents of the form $G\{\phi, M \Rightarrow N, \phi\}$ are derivable in $\mathbf{NS}_\mathsf{k} \cup X$.*

***Proof*** By induction on $\phi$. □

The structural rules of standard sequent calculi are adapted to the nested structure: they can be applied to any world-sequent (see rules $\mathsf{WL}$, $\mathsf{WR}$ and $\mathsf{CL}$, $\mathsf{CR}$ of Fig. 11.6). In addition to them we also have rules manipulating the nested sequent structure, namely the external structural rules of *external weakening* ($\mathsf{EW}$), *merge* ($\mathsf{merge}$) and *necessitation* ($\mathsf{NR}$), see Fig. 11.6. It can be shown that all these rules are *height-preserving admissible*, i.e., that whenever their premisses are derivable with derivations of height at most $n$, then so is their conclusion.

**Lemma 11.2.14** (Admissibility of the structural rules) *For any $X \subseteq \{d, t, b, 4\}$ under the 5-closure proviso, the rules* $\mathsf{WL}$, $\mathsf{WR}$, $\mathsf{EW}$, $\mathsf{CL}$, $\mathsf{CR}$, $\mathsf{NR}$ *and* $\mathsf{merge}$ *(see Fig. 11.6) are height-preserving admissible in $\mathbf{NS}_\mathsf{k} \cup X$.*

$$\frac{G\{M \Rightarrow N\}}{G\{\alpha, M \Rightarrow N\}}\ \mathsf{WL} \qquad \frac{G\{M \Rightarrow N\}}{G\{M \Rightarrow N, \alpha\}}\ \mathsf{WR} \qquad \frac{G\{M \Rightarrow N\}}{G\{M \Rightarrow N, [P \Rightarrow Q]\}}\ \mathsf{EW}$$

$$\frac{G\{\alpha, \alpha, M \Rightarrow N]}{G\{\alpha, M \Rightarrow N\}}\ \mathsf{CL} \qquad \frac{G\{M \Rightarrow N, \alpha, \alpha\}}{G\{M \Rightarrow N, \alpha\}}\ \mathsf{CR} \qquad \frac{G\{M \Rightarrow N, [P \Rightarrow Q], [S \Rightarrow T]\}}{G\{M \Rightarrow N, [P, S \Rightarrow Q, T]\}}\ \mathsf{merge}$$

$$\frac{G}{\Rightarrow [G]}\ \mathsf{NR}$$

**Fig. 11.6** Structural rules

***Proof*** By induction on the derivation of the premiss(es). □

Using the previous lemma we can then show that the logical and special logical rules are *height-preserving invertible*, i.e., whenever their conclusion is derivable with a derivation of height at most $n$, then so are their premisses.

**Lemma 11.2.15** (Invertibility) *For any $X \subseteq \{d, t, b, 4\}$ under the 5-closure proviso, the logical, modal and special logical rules are height-preserving invertible in* $\mathbf{NS}_{\mathsf{k}} \cup X$.

***Proof*** By induction on the derivation of the premiss(es), using Lemma 11.2.14. □

As we have seen above, each of the axioms listed in Fig. 11.2 corresponds to a special logical rule (see Fig. 11.3); however, it also corresponds to a *special structural rule*, see Fig. 11.8. Each of the special structural rules mirrors the frame property corresponding to the axiom to which it is linked. For example, rule $\tilde{d}$ ensures that every node in a nested sequent can be equipped with a successor, corresponding to the seriality property linked to the axiom $\mathsf{d}$. Given the special logical rules, we then have admissibility of the corresponding special structural rules.

**Lemma 11.2.16** (Admissibility of the special structural rules) *For any $X \subseteq \{d, t, b, 4, 5\}$, let $\tilde{X}$ be the corresponding subset in $\{\tilde{d}, \tilde{t}, \tilde{b}, \tilde{4}, \tilde{5}\}$ (see Fig. 11.8). For any $X \subseteq \{d, t, b, 4, 5\}$, the rules in $\tilde{X}$ are height-preserving admissible in* $\mathbf{NS}_{\mathsf{k}} \cup X$.

***Proof*** By induction on the derivation of the premiss(es). □

### *11.2.3 Cut-Elimination, Soundness and Completeness*

One of the technical benefits of the nested sequent framework is that it can be used in purely syntactic *cut elimination* proofs. However, defining the cut rule in this framework is not entirely trivial, since we need to generalise it to the tree structure of nested sequents. As usual we have the choice between the *context-sharing* and the *context-splitting* variant of the rule. The context-sharing variant is given by:

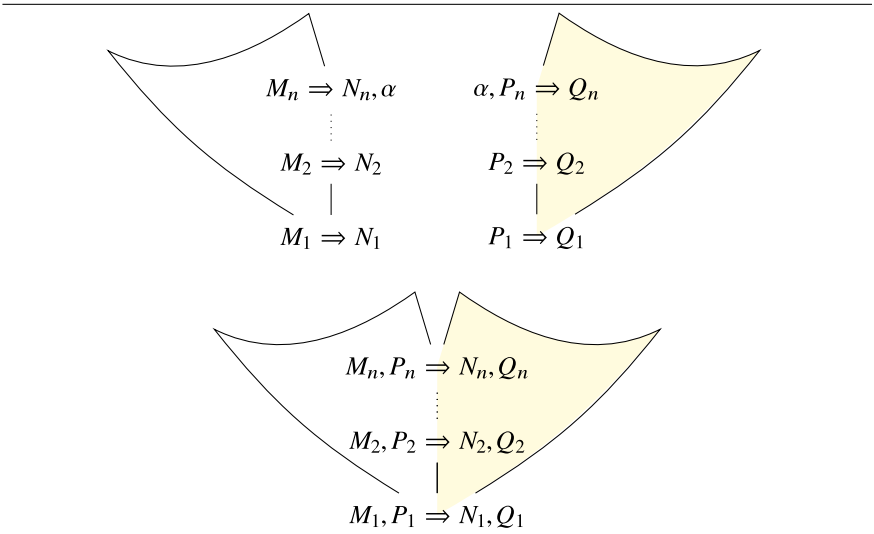



**Fig. 11.7** A visual representation of the premisses and conclusion of the (context-splitting) cut rule for nested sequents

$$\frac{G\{M \Rightarrow N, \phi\} \quad G\{\phi, M \Rightarrow N\}}{G\{M \Rightarrow N\}}\ \mathsf{cut}_{cs}$$

While cut elimination can indeed be shown syntactically for the context-sharing variant of the cut rule (see, e.g., Brünnler, 2009), this entails a certain overhead in terms of the structural rules to ensure that the context of the premisses is always the same. Hence in the following we consider the context-splitting variant. The intuition is that considering the rule bottom-up from conclusion to premisses, the nested sequent is split along the branch from the root to the node where the cut formula is introduced. Dually, considering the rule from premisses to conclusion, the nested sequents corresponding to the premisses are *merged* or *zipped together* along the branch from their roots to the node containing the cut formula. The graphical intuition for this is shown in Fig. 11.7. Formally, this is captured as follows (recall that we abuse notation and write, e.g., $M_1 \Rightarrow N_1$ for $M_1 \Rightarrow N_1, [G_1], \ldots, [G_m]$):

**Definition 11.2.17** For two nested sequent contexts

$$G\{\,\} = M_1 \Rightarrow N_1, [\ldots [M_n \Rightarrow N_n, [\{\,\}]] \ldots]$$
$$H\{\,\} = P_1 \Rightarrow Q_1, [\ldots [P_n \Rightarrow Q_n, [\{\,\}]] \ldots]$$

the *merge* is the following nested sequent context:

$$G \oplus H\{\,\} := M_1, P_1 \Rightarrow N_1, Q_1, [\ldots [M_n, P_n \Rightarrow N_n, Q_n, [\{\,\}]] \ldots] .$$

$$\frac{G\{M \Rightarrow N, [\Rightarrow]\}}{G\{M \Rightarrow N\}}\ \tilde{d} \qquad \frac{G\{M \Rightarrow N, [P \Rightarrow Q]\}}{G\{M, P \Rightarrow N, Q\}}\ \tilde{t}$$

$$\frac{G\{M \Rightarrow N, [P \Rightarrow Q]\}}{G\{M \Rightarrow N, [\ \Rightarrow [P \Rightarrow Q]]\}}\ \tilde{4} \qquad \frac{G\{M \Rightarrow N, [S \Rightarrow T, [P \Rightarrow Q]]\}}{G\{M, P \Rightarrow N, Q, [S \Rightarrow T]\}}\ \tilde{b}$$

$$\frac{G\{M \Rightarrow N, [S \Rightarrow T, [P \Rightarrow Q]]\}}{G\{M \Rightarrow N, [S \Rightarrow T, P \Rightarrow Q]\}}\ \tilde{5}$$

**Fig. 11.8** The special structural rules for extensions of modal logic K

Intuitively, in $G \oplus H\{\,\}$ the two nested sequent contexts are "zipped together" along the path from root to the context symbol $\{\,\}$. In particular, this means that the context symbol $\{\,\}$ occurs at the same depth both in $G$ and $H$.

The *context-splitting cut rule* in the nested sequent framework then is the following rule, where the context has the same depth in $G\{\,\}$ and $H\{\,\}$:

$$\frac{G\{M \Rightarrow N, \phi\} \quad H\{\phi, P \Rightarrow Q\}}{G \oplus H\{M, P \Rightarrow N, Q\}}\ \text{cut}$$

The cut elimination result is then shown as follows.

**Lemma 11.2.18** (Cut elimination) *Let $G\{M \Rightarrow N, \phi\}$ and $H\{\phi, P \Rightarrow Q\}$ be two nested sequents where the context symbol $\{\,\}$ occurs at the same depth both in $G$ and $H$. If*

$$\frac{\begin{array}{cc}\vdots\ d_1 & \vdots\ d_2\\ G\{M \Rightarrow N, \phi\} & H\{\phi, P \Rightarrow Q\}]\end{array}}{G \oplus H\{M, P \Rightarrow N, Q\}}\ \text{cut}$$

*and $d_1$ and $d_2$ do not contain any other application of the cut-rule, then we can construct a derivation of $G \oplus H\{M, P \Rightarrow N, Q\}$ with no application of the cut-rule.*

***Proof*** The proof is by induction on the complexity of the cut-formula $\phi$, with subinduction on the sum of the heights of the derivations of the premisses of the cut-rule. The details of the proof can be found in Poggiolesi (2008b, Sect. 7); here we just illustrate some significant cases. Suppose the cut-formula $\phi$ is of the form $\Box\epsilon$ and it has been introduced in the left premiss by the rule $\Box R$ and in the right premiss by the rule $\Box L$. Hence we have the following situation:

$$\frac{\dfrac{G\{M \Rightarrow N, [\Rightarrow \epsilon\,]\}}{G\{M \Rightarrow N, \Box\epsilon\,\}}\ \Box R \qquad \dfrac{H\{\Box\epsilon\,, P \Rightarrow Q, [\epsilon\,, S \Rightarrow T]\}}{H\{\Box\epsilon\,, P \Rightarrow Q, [S \Rightarrow T]\}}\ \Box L}{G \oplus H\{M, P \Rightarrow N, Q, [S \Rightarrow T]\}}\ \text{cut}$$

We reduce to

$$\dfrac{\dfrac{G\{M \Rightarrow N, [\Rightarrow \epsilon\,]\} \qquad \dfrac{G\{M \Rightarrow N, \Box\epsilon\,\} \quad H\{\Box\epsilon\,, P \Rightarrow Q, [\epsilon\,, S \Rightarrow T]\}}{G \oplus H\{M, P \Rightarrow N, Q, [\epsilon\,, S \Rightarrow T]\}}\ \mathsf{cut}}{G \oplus G \oplus H\{M, M, P \Rightarrow N, N, Q, [S \Rightarrow T]\}}\ \mathsf{cut}}{G \oplus H\{M, P \Rightarrow N, Q, [S \Rightarrow T]\}}\ X$$

where the first cut is eliminable by induction on the height of the proof, whilst the second cut is eliminable by induction on the complexity of the cut-formula. The $X$ denotes several applications of internal contraction rules and merge.

Several other interesting cases of cut emerge whenever the cut-formula $\phi$ is still of the form $\Box\epsilon$, the left premiss has been introduced by the rule $\Box R$ and the right premiss has been introduced by one of the special logical rules $Y$. In the cases where $Y$ is either $t, d$ or $b$, the strategy is the same: in order to reduce, we need to use the corresponding admissible special structural rule $\tilde{Y}$, see Lemma 11.2.16, on the nested sequent above the left premiss of the cut. As an example, we show the case of the rule $t$.

$$\dfrac{\dfrac{G\{M \Rightarrow N, [\Rightarrow \epsilon\,]\}}{G\{M \Rightarrow N, \Box\epsilon\,\}}\ \Box R \qquad \dfrac{H\{\Box\epsilon\,, \epsilon\,, P \Rightarrow Q\}}{H\{\Box\epsilon\,, P \Rightarrow Q\}}\ t}{G \oplus H\{M, P \Rightarrow N, Q\}}\ \mathsf{cut}$$

We reduce to

$$\dfrac{\dfrac{G\{M \Rightarrow N, [\Rightarrow \epsilon\,]\}}{G\{M \Rightarrow N, \epsilon\,\}}\ \tilde{t} \qquad \dfrac{G\{M \Rightarrow N, \Box\epsilon\,\} \quad H\{\Box\epsilon\,, \epsilon\,, P \Rightarrow Q\}}{G \oplus H\{\epsilon\,, M, P \Rightarrow N, Q\}}\ \mathsf{cut}}{G \oplus G \oplus H\{M, M, P \Rightarrow N, N, Q\}}\ \mathsf{cut}}{G \oplus H\{M, P \Rightarrow N, Q\}}\ X$$

where the first cut is eliminable by induction on the height of the proof, whilst the second cut is eliminable by induction on the complexity of the cut-formula. The $X$ denotes several applications of internal contraction rules and merge.

In the cases where the special logical rule used to introduce the right premiss of the cut-rule is either 4 or 5, then the strategy is analogous to the one above except for the fact that one should further distinguish cases on what may have introduced the nested sequent above the right premiss of the cut-rule. □

**Theorem 11.2.19** *For any $X \subseteq \{d, t, b, 4\}$ under the 5-closure proviso, every derivation in $\mathbf{NS}_{\mathsf{k}} \cup X \cup \{\mathsf{cut}\}$ can be effectively transformed into a derivation $d'$ in $\mathbf{NS}_{\mathsf{k}} \cup X$, i.e., where there is no application of the cut-rule.*

***Proof*** It follows from the previous Lemma by induction on the number of cuts. □

We can now close the section by establishing soundness and completeness for the nested calculi.

**Theorem 11.2.20** *For any $X \subseteq \{d, t, b, 4\}$ under the 5-closure proviso, the nested calculus $\mathbf{NS}_{\mathsf{k}} \cup X$ is sound and complete with respect to the corresponding Hilbert system.*

***Proof*** The soundness proof is by induction on the height of derivations in Hilbert systems and can be found in Poggiolesi (2010b, Sect. 6.3). The completeness proof consists in showing that we can construct in nested sequent calculi derivations of the axioms and the inference rules that compose the corresponding Hilbert system. Here we show the derivation of the distribution axiom; the others derivations can be found in Poggiolesi (2010b, Sect. 6.3).

$$\dfrac{\dfrac{\dfrac{\dfrac{\dfrac{\dfrac{\Box(\phi \to \epsilon), \Box\phi \Rightarrow [\phi \Rightarrow \phi, \epsilon] \quad \Box(\phi \to \epsilon), \Box\phi \Rightarrow [\phi, \epsilon \Rightarrow \epsilon]}{\Box(\phi \to \epsilon), \Box\phi \Rightarrow [\phi \to \epsilon, \phi \Rightarrow \epsilon]}\,{\to}L}{\Box(\phi \to \epsilon), \Box\phi \Rightarrow [\phi \Rightarrow \epsilon]}\,\Box L}{\Box(\phi \to \epsilon), \Box\phi \Rightarrow [\Rightarrow \epsilon]}\,\Box L}{\Box(\phi \to \epsilon), \Box\phi \Rightarrow \Box\epsilon}\,\Box R}{\Box(\phi \to \epsilon) \Rightarrow \Box\phi \to \Box\epsilon}\,{\to}R}{\Rightarrow \Box(\phi \to \epsilon) \to \Box\phi \to \Box\epsilon}\,{\to}R$$

Note that in the completeness proof, use of a nested form of the *cut rule* to simulate applications of modus ponens in the Hilbert-system is required. There also exists a completeness proof based on the construction of counter-models that can be found in Brünnler (2006). We will look at this proof in detail in Sect. 11.3.1. □

The nested calculi so far introduced enjoy several interesting properties that we hope to have underlined; however they also display a defect which is the lack of *modularity*.[7] The nested calculi presented are not modular because the axiom 5 is not systematically reflected by the rule 5. As shown by Marin and Straßburger (2014), this lack of modularity can be repaired by giving up the logical variant of the nested calculi. Indeed, if we consider nested sequent calculi where: (i) the contraction rules CL and CR are added (see Fig. 11.6), (ii) in the rule $\Box L$ as well as in the special logical rules, the formula $\Box\phi$ which is not the main formula of the rule is erased, (ii) for each modal axiom, the corresponding special logical and structural rule are taken as primitive rules, then we have a more general version of the rules 5 and $\tilde{5}$ which allow the calculi to be modular.

## 11.3 Applications

In this section we illustrate three applications of the nested sequent method; in particular, we will dwell on countermodels, decidability and interpolation results.

[7] Consider a Hilbert system $S'$ which is obtained by adding an axiom $A$ to the Hilbert system $H$. The modularity property demands the Gentzen system $G'$ for the system $S'$ to be obtained from the system $G$ for the Hilbert system $S$ by adding $n$ rules that systematically reflect at the proof-theoretical level the axiom $A$.

### 11.3.1 Countermodel Construction

In this section we will focus on the completeness of nested calculi via the construction of countermodels. More precisely, we show the strategy of this proof in the case of the nested calculus $\mathbf{NS_k}$. Note that the completeness proof via countermodels can be proved for all the nested calculi $\mathbf{NS_k} \cup X$, for any $X \subseteq \{d, t, b, 4\}$ under the 5-closure proviso; in these cases, the proof is analogous to that for the nested calculus $\mathbf{NS_k}$, except that properties of the accessibility relation in Kripke frames need to be taken into account as well, see Brünnler (2006) and Poggiolesi (2010b, Sect. 8).

For each rule $\mathcal{R}$ of $\mathbf{NS_k}$ we consider the *Kleene'd version*, i.e., the rule $\mathcal{R}^+$ that is the same as $\mathcal{R}$ except that it maintains the main formula of the conclusion in the premiss (analogously to the G3 systems in Kleene 1952). For example, if $\mathcal{R}$ is the rule $\neg R$, then $\mathcal{R}^+$ will be:

$$\frac{G\{\phi, M \Rightarrow N, \neg\phi\}}{G\{M \Rightarrow N, \neg\phi\}} \; \neg^+ R$$

Let $\mathbf{NS_k}^+$ be the calculus obtained by replacing each rule $\mathcal{R}$ in $\mathbf{NS_k}$ with the corresponding rule $\mathcal{R}^+$. It is straightforward to show that the calculi $\mathbf{NS_k}^+$ and $\mathbf{NS_k}$ are equivalent (see Brünnler, 2006). Thus we can essentially consider nested sequents based on sets instead of multisets:

**Definition 11.3.1** The *set nested sequent* of the nested sequent $M \Rightarrow N, [G_1], \ldots, [G_n]$ is the nested sequent $P \Rightarrow Q, [H_1], \ldots, [H_n]$, where $P$ and $Q$ are the underlying sets of $M$ and $N$, respectively, and $H_1, \ldots, H_n$ are the set nested sequents of $G_1, \ldots, G_n$, respectively. Clearly the set nested sequent of a nested sequent is still a nested sequent since a set is a multiset.

**Definition 11.3.2** A leaf of a nested sequent (thinking of the nested sequent as a tree-frame of Kripke semantics) is *cyclic* if in its branch there exists a second sequent that contains the same set of formulae.

**Definition 11.3.3** A sequent occurring in a nested sequent is *saturated* for a nested sequent calculus $\mathbf{Y}$ if no rule of that calculus can be applied to one of its formulae to obtain a formula which is not already present in that sequent. A nested sequent is finished for a nested sequent calculus $\mathbf{Y}$ if all sequents that compose it are finished or cyclic.

**Definition 11.3.4** We define a procedure $prove(G, \mathbf{NS_k}^+)$, which takes a nested sequent $G$ and the calculus $\mathbf{NS_k}^+$, and builds a derivation tree for $G$ by applying rules from that calculus to non-initial and unfinished derivation leaves in a bottom-up fashion, as follows:

1. keep applying all the rules of $\mathbf{NS_k}^+$ which are not the rule $(\Box K)^+$;
2. wherever possible, apply the rule $(\Box K)^+$ once.

Repeat this operation until each non-initial derivation leaf of the nested sequent $G$ is finished. If $prove(G, \mathbf{NS}^+_{\mathsf{k}})$ terminates and all derivation leaves are initial then it succeeds; otherwise, i.e., if it terminates and there is a non-initial derivation leaf, it fails.

**Definition 11.3.5** The *set of subformulae* of a nested sequent $G$, denoted *sf(G)*, is the set of all subformulae of all formulae that compose all sequents that belong to the nested sequent.

**Lemma 11.3.6** *For all nested sequents $G$, the procedure $prove(G, \mathbf{NS}^+_{\mathsf{k}})$ terminates after at most $2^{(|sf(G)|)}$ iterations.*

***Proof*** Consider a sequence of nested sequents along a given branch of the derivation starting from the root. None of the rules that we can apply in accordance with step 1 creates new sequents in the nested sequents, but each of them causes the set of formulae of some sequent belonging to the nested sequent to strictly grow. By the subformula property, only $|sf(G)|$ many formulae can occur in each side of a sequent, so step 1 terminates. If after step 1 there is an unfinished leaf in a nested sequent, then the size of the nested sequent strictly grows in step 2. Since there are only $2^{|sf(G)|}$ different sets of formulae that can occur, each unfinished nested sequent leaf has to be cyclic eventually. □

In the following definition we exploit the strong analogy between nested sequents and tree-frames of Kripke semantics (see Definition 11.2.4).

**Definition 11.3.7** A nested sequent $H$ is an *immediate subtree* of a nested sequent $G$ if $G$ is of the form $M \Rightarrow N, [H], [H_1], \ldots, [H_n]$. It is a *proper subtree* if it is an immediate subtree either of $G$ or of a proper subtree of $G$, and it is a *subtree* if it is either a proper subtree of $G$ or $G = H$. The *set of all subtrees* of $G$ is denoted by *st(G)*.

**Lemma 11.3.8** *For all nested sequents $G$, if $prove(G, \mathbf{NS}^+_{\mathsf{k}})$ fails, then there is a countermodel for $G$.*

***Proof*** We give a sketch of the proof by indicating how to construct the countermodel for $G$ after $prove(G, \mathbf{NS}^+_{\mathsf{k}})$ has failed. A Kripke model (see Definition 11.2.5) is composed of (i) a set of worlds $W$, (ii) an accessibility relation amongst worlds $\rightarrow$ and (iii) an evaluation function $v$ mapping formulae and worlds into truth values. Here is how we construct each of the elements of the Kripke model $\mathfrak{M}$ intended to be a countermodel for $G$.

(i) Let $G^*$ be the set nested sequent of the non-initial nested sequent obtained by the failure of $prove(G, \mathbf{NS}^+_{\mathsf{k}})$. Let $\Upsilon$ be the set of all cyclic leaves in $G^*$. Then in the model $\mathfrak{M}$ the set of worlds is obtained by considering the set of all subtrees of $G^*$ minus $\Upsilon$, i.e., $W = st(G^*) \setminus \Upsilon$. Let $f : \Upsilon \rightarrow W$ be some function which maps a cyclic leaf to a nested sequent in $W$ whose root contains the same set of formulae, and extend $f$ to $st(G^*)$ by the identity on $W$.

(ii) The accessibility relation on $W$ is defined in the following way. For all $G$ and $H$, $G \to H$ if, and only if, either $H$ is an immediate subtree of $G$, or $G$ has an immediate subtree $H_1$, such that $H_1 \in \Upsilon$ and $f(H_1) = H$.

(iii) Finally the evaluation $v$ from propositional atoms and nested sequents to truth values is defined in the following way. An atom $p$ is true in a world-nested sequent $G$ if it occurs on the left side of a sequent belonging to the nested sequent. It is false, otherwise.

The rest of the proof aims at showing that the model $\mathfrak{M}$ is a countermodel for $G$, see Brünnler (2006) and Poggiolesi (2010b, pp. 173–174). □

The definition of validity of a nested sequent is tedious but it is a quite straightforward generalization of the notion of validity of a sequent. Its definition can be found in Poggiolesi (2010b, pp. 165–166).

**Theorem 11.3.9** *For all nested sequents G, if G is valid with respect to the class of all frames, then the nested sequent calculus* $\mathbf{NS_k}$ *is such that* $\vdash G$ *in* $\mathbf{NS_k}$.

***Proof*** It follows from the previous Lemma and the fact that the calculi $\mathbf{NS_k}$ and $\mathbf{NS_k}^+$ are equivalent. □

### *11.3.2 Decidability*

Another notable feature of nested calculi is that most of them are decidable: given a nested sequent $G$, there is an algorithm determining whether $G$ is provable or not. We dedicate this section to a brief presentation of decidability in nested calculi[8]; as we have done in the previous sections, for our purposes we will focus on the calculus $\mathbf{NS_k}$; however decidability can be proved for (most of)[9] the nested calculi $\mathbf{NS_k} \cup X$, for any $X \subseteq \{d, t, b, 4\}$ under the 5-closure proviso; only in the case of the system for S4 a certain attention needs to be dedicated to the interaction between the rules 4 and the rule $\Box K$ which might create endless chain of sequents: however, thanks to the use of the rule $t$ the problem is easily fixed, see Poggiolesi (2010b, p. 161).

We start by noticing that in a calculus enjoying the cut-elimination theorem, the *subformula property* - in any of its rules the formulae occurring in the premiss(es) are subformulae of the formulae occurring in the conclusion—and contraction-admissibility, any source of potential non-terminating proof-search seems to be avoided. Although the calculus $\mathbf{NS_k}$ enjoys all these features, because of the rule $\Box L$ where the same formula $\Box\phi$ occurs in the conclusion as well as in the premiss, we still have potentially endless applications. In order to overcome this problem, a solution is to work with *minimal derivations* namely derivations where shortening

[8] Note that decidability for modal logic is a well-known result since the 1960s, e.g., see Segerberg (1968).

[9] The nested calculi for which a decidability proof is developed are $\mathbf{NS_k}$, $\mathbf{NS_{kt}}$, $\mathbf{NS_{kd}}$, $\mathbf{NS_{kb}}$, $\mathbf{NS_{kbt}}$, $\mathbf{NS_{S4}}$, $\mathbf{NS_{S5}}$

are not possible in the sense that every subderivation is of minimal height. Suppose that the rule $\Box L$ is used twice on the same pair of sequents in a minimal derivations, then we can show that such derivation is no longer minimal and hence that the rule $\Box L$ can at most be used once. Thanks to this device, we can prove that the calculus $\mathbf{NS_k}$ allows for terminating proof-search.

As shown in Poggiolesi (2010b), in the other nested calculi the problem raised by the rule $\Box L$ also arises with the special logical rules. However, minimal derivations represent again the solution to such a problem: in these derivations special logical rules cannot be applied more than once on the same formula or pair of sequents.

**Lemma 11.3.10** *The rule $\Box L$ permutes down with respect to the propositional rules. It also permutes with instances of the rule $\Box R$ in the case where their auxiliary formulae, $\phi$ and $\Box\phi$, respectively, are not active in the sequent where the auxiliary formula of $\Box R$ occurs.*

***Proof*** The proof is straightforward. □

**Corollary 11.3.11** *In a minimal derivation in $\mathbf{NS_k}$, the rule $\Box L$ cannot be applied more than once on the same pair of formulas belonging to the same pairs of sequents of any branch.*

***Proof*** Let us suppose we have a minimal derivation where the rule $\Box L$ has been applied twice on the same pair of sequents,

$$\dfrac{\dfrac{G\{\Box\phi, M' \Rightarrow N', [\phi, P' \Rightarrow Q']\}}{G\{\Box\phi, M' \Rightarrow N', [P' \Rightarrow Q']\}}\,\Box L}{\vdots}$$
$$\dfrac{G\{\Box\phi, M \Rightarrow N, [\phi, P \Rightarrow Q]\}}{G\{\Box\phi, M \Rightarrow N, [P \Rightarrow Q]\}}\,\Box L$$

By permuting down $\Box L$ with respect to the steps in the dotted part of the derivation, we obtain a derivation of the same height ending with

$$\dfrac{\dfrac{G\{\Box\phi, M \Rightarrow N, [\phi, \phi, P \Rightarrow Q]\}}{G\{\Box\phi, M \Rightarrow N, [\phi, P \Rightarrow Q]\}}\,\Box L}{G\{\Box\phi, M \Rightarrow N, [P \Rightarrow Q]\}}\,\Box L$$

By applying the height-preserving admissible rule $\mathsf{CL}$ to the two occurrences of the formula $\phi$ in place of the upper $\Box L$, we obtain a shortened derivation, contrary to the assumption of minimality. □

We can finally prove that the modal logic **K** is decidable by means of showing effective bounds on proof search in the calculus $\mathbf{NS_k}$.

**Theorem 11.3.12** *The calculus $\mathbf{NS_k}$ allows terminating proof search.*

***Proof*** Place a nested sequent $G$, for which we are looking for a proof search, at the root of the procedure. Apply first the propositional rules and then the modal rules. The propositional rules reduce the complexity of the nested sequent. The rule $\Box R$ removes the modal constant $\Box$, and adds a new sequent, the rule $\Box L$ increases the complexity. However, by the corollary above, the rule $\Box L$ cannot be applied more than once to the same pair of sequents with the same principal formula. Therefore, the number of applications of the rule $\Box L$ is bounded by the number of subformulae of the conclusion and the number of sequents that may appear in the derivation. The latter, in its turn, is bounded by the number of sequents belonging to the nested sequents to prove, and the sequents that can be introduced by applications of the rule $\Box R$. □

### *11.3.3 Interpolation*

A logic $\mathsf{L}$ is said to enjoy the *Craig Interpolation Property*, for short *CIP*, if whenever, $A \vdash_{\mathsf{L}} B$, there exists a formula $C$, called the *interpolant*, that stands in between $A$ and $B$ in terms of derivability, namely such that $A \vdash_{\mathsf{L}} C$, $C \vdash_{\mathsf{L}} B$ and $C$ contains only propositional variables common to $A$ and $B$. Since in many logics where the CIP is proved, it is possible to define an implication enjoying modus ponens and the deduction theorem, a more common formulation of CIP uses $\vdash_{\mathsf{L}} A \to B$ instead of $A \vdash_{\mathsf{L}} B$. The CIP is a deep feature of a logic: it has notable applications in computer science (e.g., consistency proofs, model checking), it is connected to Beth definability (e.g., see Hodges, 1997) and it can be translated to the amalgamation property of algebraic varieties (e.g., see Day & Jaroslav, 1984). However the CIP might fail even in some reasonable fundamental logics, such as intuitionistic logic with constant domains, see Mints et al. (2013).

The topic of this section is the Craig Interpolation Property for the main systems of modal logic. There are several different ways to prove this result (e.g., see Fitting, 1983); in these pages the focus goes on a constructive method for proving interpolants for modal logic, namely using induction on derivations in nested calculi. The first research in this direction has been initiated by Bílková (2011) for the stronger *uniform* interpolation property. An extensive body of work was further developed starting from Fitting and Kuznets (2015) and Kuznets (2018). In Lyon et al. (2020) the technique adopted by Fitting and Kuznets has been developed in a purely syntactical variant independent of the semantics of the corresponding logics and applied to the realm of logics with converser modalities including tense bi-intuitionistic logics. In what follows, we will mainly rely on Fitting and Kuznets's results; however, given the length and complexity of these results, we will not dwell on the details but only sketch their main features and key-points.

In order to understand how the CIP can be proved via nested calculi, let us start from how it is proved in a simpler framework, namely that of sequent calculi for classical logic. Consider then a theorem of classical logic of the form $A \to B$; since it is a theorem, there will be a derivation in the sequent calculus **Gcl** of the sequent

$A \Rightarrow B$ and the derivation will possibly involve other sequents of the form $M \Rightarrow N$. For constructing the interpolant $C$, we do not look at the sequents $A \Rightarrow B$ or $M \Rightarrow N$ themselves, but to their *split* version, namely:

$$A; \Rightarrow; B \qquad M_1; M_2 \Rightarrow N_1; N_2$$

The need of splitting sequents comes from the fact that certain rules, e.g., for the propositional negation and implication, move formulae around between the antecedent and consequent and thus disqualify them from being the stable left and right side to be interpolated. Because of the split of the sequents, it is not entirely correct to say that Craig interpolation is proved by induction on the sequent derivation. In reality, one finds a suitable generalization of CIP and proves that by induction on the split version of the sequent derivation. Formally, it is proved that, for any derivable sequent $M \Rightarrow N$ and for any partition of $M = M_1 \bigcup M_2$ and $N = N_1 \bigcup N_2$ into two parts each, there exists a formula $C$ such that:

- $\vdash_{\mathsf{CL}} M_1 \Rightarrow N_1, C$,
- $\vdash_{\mathsf{CL}} C, M_2 \Rightarrow N_2$,
- each atomic proposition occurring in $C$ occurs both in $M_1 \Rightarrow N_1$ and $M_2 \Rightarrow N_2$.

The fact that $C$ is an interpolant is ensured by the standard interpretation of sequents. Indeed we have (see Sect. 11.2):

$$(M_1, M_2 \Rightarrow N_1, N_2)^{\Delta} = \bigwedge M_1 \wedge \bigwedge M_2 \rightarrow \bigvee N_1 \wedge \bigwedge N_2$$

that is equivalent to

$$\neg(\bigwedge M_1 \rightarrow \bigvee N_1) \rightarrow (\bigwedge M_2 \rightarrow \bigvee N_2)$$

The idea of split sequents can be found in Maehara (1961) and is a key feature of what is called the *Maehara method* for constructing interpolants via the Gentzen system. Since such a method is extended and employed by Fitting and Kuznets for nested calculi for modal logics, let us briefly describe it. Given a cut-free sequent calculus for the logic $\mathsf{CL}$, and a theorem of the form $A \rightarrow B$ of $\mathsf{CL}$:

1. find a sequent derivation of $A \Rightarrow B$,
2. split the endsequent to be $A; \Rightarrow; B$,
3. propagate this split from the endsequent to the leaves turning the derivation into a split derivation,
4. find interpolants for each split initial sequent in a leaf,
5. propagate interpolants from the leaves to the endsequent using rules for transforming interpolants for each rule.

Such method comes with tables presenting split initial sequents, split sequent rules, but also with sufficient interpolant transformations. We will not present these tables here, they can be found in Kuznets (2018, 1377–1378).

Let us now move to modal logic and to nested calculi. At the first glance it might seem that the Maehara method can be straightforwardly applied to this extension of classical logic given that nested calculi are cut-free and they only contain logical rules. Unfortunately, this is not so obvious for the following reason. Consider a quite simple nested sequent of the following form:

$$M \Rightarrow N, [P \Rightarrow Q], [S \Rightarrow T]$$

The splitting should be performed by splitting each of the sequent components:

$$M_1, M_2 \Rightarrow N_1, N_2, [P_1, P_2 \Rightarrow Q_1, Q_2], [S_1, S_2 \Rightarrow T_1, T_2]$$

According to the standard interpretation of a nested sequent (see Definition 11.2.9), this nested sequent corresponds to the following formula:

$$(\bigwedge M_1 \wedge \bigwedge M_2 \rightarrow \bigvee N_1 \vee \bigvee N_2) \vee \Box(\bigwedge P_1 \wedge \bigwedge P_2 \rightarrow \bigvee Q_1 \vee \bigvee Q_2) \vee \Box(\bigwedge S_1 \wedge \bigwedge S_2 \Rightarrow \bigvee T_1 \vee \bigvee T_2)$$

As far as it is known, it seems there is no way to find an equivalent syntactical representation of this formula as an implication having $M_1, N_1, P_1, Q_1, S_1, T_1$ in its antecedent and $M_2, N_2, P_2, Q_2, S_2, T_2$ in its consequent. Does this mean that the Maehara method cannot be applied to modal logics (via nested sequents) or is there another way of splitting a nested sequent? Thanks to a careful use of semantic methods, Fitting and Kuznets propose a novel and fine-grained way to split a nested sequent.

**Definition 11.3.13** Let $\mathfrak{M} = (W, R, v)$ be a Kripke model with the set of worlds $W \neq \emptyset$, the accessibility relation $R$ amongst worlds, and the propositional valuation $v$. We define $\mathfrak{M}$-*decorated nested objects* $G^{\Psi}$ as a nested sequent with each sequent node decorated by a world from W in such a way that the world decorating a child of a given sequent node is R-accessible from the world decorating the sequent node itself.

**Definition 11.3.14** Let $\mathfrak{M} = (W, R, v)$ be a Kripke model. An $\mathfrak{M}$-decorated nested object $G^{\Psi}$ is *true* if at least one formula from $G^{\Psi}$ holds at the world of $W$ that decorates the sequent node of this formula. We write $\models_{\mathfrak{M}} G^{\Psi}$ to denote that $G^{\Psi}$ is true in the model $\mathfrak{M}$.

The crucial idea behind this definition is that decorations help us to look at nested sequents as disjunctions of member formulae, similar to the standard interpretation of normal sequents. However, while in the case of standard sequents, all formulae are evaluated at the same world of the Kripke model, in the nested case, formulae from different sequent nodes are evaluated at different worlds of the Kripke model.

Consider again the nested sequent above, namely:

$$M_1, M_2 \Rightarrow N_1, N_2, [P_1, P_2 \Rightarrow Q_1, Q_2], [S_1, S_2 \Rightarrow T_1, T_2]$$

Then appropriately decorate this nested sequent by the three worlds $w$, $w_1$, $w_2$, such that $wRw_1$ and $wRw_2$. We can finally define the split of the nested sequent in the following adequate way:

$$\models_{\mathfrak{M}} M_1 \Rightarrow N_1, [P_1 \Rightarrow Q_1], [S_1 \Rightarrow T_1] \text{ implies } \models_{\mathfrak{M}} M_2 \Rightarrow N_2, [P_2 \Rightarrow Q_2], [S_2 \Rightarrow T_2]$$

We have defined how to split a nested sequent and how to evaluate it by decorating sequent nodes. It remains to define interpolants, which will be intermediaries of implications of the types above. However, implications of this sort are not purely syntactic anymore: they involve sequents as well as multiple worlds where these sequents are evaluated. As a consequence, interpolants will have a similar form: they will not be mere formulae or mere sequents, but *generalized nested sequents*, namely a mix of nested sequents and worlds, with a particular tree-structure analogous to that of the nested sequents they are the intermediaries for. Generalized nested sequents are defined in Fitting and Kuznets (2015, p. 284).

In order to prove the CIP via nested calculi, in Fitting and Kuznet (2015, p. 284) as well as in Kuznets (2018) (although with different notation and names), tables for splitting initial nested sequents, splitting nested sequent rules, but also tables with sufficient interpolant transformations are given. We are forced to omit all these tables together with the definitions and properties of generalized nested sequents because they are all too long and laborious to be briefly summed up here. However they represent the other key-ingredient to apply the Mahera method to the case of modal logic.

**Theorem 11.3.15** *For any $X \subseteq \{d, t, b, 4\}$ under the 5-closure proviso, the nested calculus $\mathbf{NS}_{\mathsf{k}} \cup X$ allow to prove that the corresponding modal logic enjoys the Craig interpolation property.*

## 11.4 Other Logics

In line with the predominant historical development of the nested sequent formalism in this article so far we concentrated on the application of this formalism to modal logics, and in particular to normal modal logics. Due to its expressiveness and flexibility the formalism however also has been applied to other logics. In this section we present a brief overview over the range of such applications. For reasons of space we only sketch some examples of calculi and refer to the cited literature for the details. We also do not discuss the applications to substructural logics via linear nested sequents in detail and refer the interested reader instead to Guerrini et al. (1998) and Lellmann et al. (2017).

### 11.4.1 Intuitionistic and First Order Logics

Due to the close connection between normal modal logics and intuitionistic or intermediate logics via Kripke or possible world semantics it is rather natural that the nested sequent formalism was also applied to intuitionistic logic. This has been done first in Postniece (2009, 2010), based on the previously developed calculi from Goré et al. (2008). Perhaps owing to previous works on display calculi, these works consider calculi for *bi-intuitionistic logic*, an extension of intuitionistic logic with a *co-implication*. Restricting to the language without the co-implication however immediately yields a nested sequent calculus for intuitionistic logic, hence for the sake of presentation here we restrict ourselves to this case. In line with the idea that different components in the nested sequent correspond to different worlds in the Kripke model, the nesting operator in the framework of intuitionistic logic now corresponds to the implication instead of the modal box. Thus, the formula translation of a nested sequent in this setting is obtained by deleting the boxes from the modal formula interpretation of Sect. 11.2. For example, the nested sequent $(M \Rightarrow N, [P_1 \Rightarrow Q_1], \ldots, [P_n \Rightarrow Q_n])$ where $N, Q_1, \ldots, Q_n$ only contain formulae is given by

$$\bigwedge M \to \bigvee N \vee \left(\bigwedge P_1 \to \bigvee Q_1\right) \vee \cdots \vee \left(\bigwedge P_n \to \bigvee Q_n\right) .$$

Note that the nesting operator is interpreted by the intuitionistic implication, hence several layers of nestings correspond to nested implications. This corresponds intuitively to the fact that in Kripke semantics for intuitionistic logic the implication is evaluated via the successor worlds. Hence, again, the different nodes of a nested sequent correspond to worlds in a Kripke model. The logical rules for the implication then are (for the sake of clarity in the form where formula contraction is not absorbed into the rule set, i.e., the principal formula is not repeated in the premisses):

$$\frac{G\{M, \epsilon \Rightarrow N\} \quad G\{M \Rightarrow N, \phi\}}{G\{M, \phi \to \epsilon \Rightarrow N\}} \to_L \qquad \frac{G\{M \Rightarrow N, [\phi \Rightarrow \epsilon\,]\}}{G\{M \Rightarrow N, \phi \to \epsilon\}} \to_R$$

The system of Postniece (2009) contains further structural rules to capture the additional properties of Kripke models for intuitionistic logic, but it was shown later in Fitting (2014) that it suffices to add a slight reformulation of one of these rules, sometimes called the *lift rule*:

$$\frac{G\{M \Rightarrow N, [P, \phi \Rightarrow Q]\}}{G\{M, \phi \Rightarrow N, [P \Rightarrow Q]\}} \text{ lift}$$

Intuitively, this rule captures the condition that valuations in Kripke models for intuitionistic logic need to be *persistent*, resulting in the fact that truth of a formula is preserved when moving to successor worlds. Completeness of the full calculi was shown by simulating previously existing so-called *shallow sequent systems* in Postniece (2009) (see *op. cit.* for the details) and by syntactic cut elimination in Postniece

(2010). For the intuitionistic fragments it was also shown by correspondence to prefixed tableaux in Fitting (2014), by simulating Maehara's standard multi-conclusion sequent calculus in Lellmann (2015), and by translating from a labelled sequent calculus in Lyon (2020a).

The similarity between the truth conditions for the intuitionistic implication and the universal quantifier was further exploited in Fitting (2014) to obtain nested sequent calculi for *first-order intuitionistic logic*, both in the constant-domain and non-constant domain formulations. In particular the nested sequent system for constant-domain intuitionistic logic is rather elegant, since only the "standard" quantifier rules need to be added. Writing $b$ for a *parameter*, i.e., one of a number of special variables for instantiating quantified variables, in the case of the universal quantifier these are:

$$\frac{G\{M, \forall x\, \phi(x), \phi(b) \Rightarrow N\}}{G\{M, \forall x\, \phi(x) \Rightarrow N\}}\ \forall_L \quad b \text{ arbitrary} \qquad \frac{G\{M \Rightarrow N, \forall x\, \phi(x), \phi(b)\}}{G\{M \Rightarrow N, \forall x\, \phi(x)\}}\ \forall_R \quad b \text{ not in conclusion}$$

The fact that the nested sequents in the conclusions of these rules contain all the worlds of a so-far constructed countermodel together with the restriction of the $\forall_R$ rule that $b$ does not occur in the conclusion of this nested sequent rule captures the requirement that the domains are constant across all worlds of the model. In the non-constant domain version these rules need to be adjusted and are more similar to the implication rules. Again, alternative completeness proofs can be given for the different versions by simulating a standard sequent system (Lellmann, 2015) or by translation from a labelled sequent calculus (Lyon, 2020a).

In the classical setting, the analogy between quantifiers and modalities was exploited even earlier, to obtain nested sequent calculi for first-order classical logic in Brünnler (2010). In this setting the nesting is seen as a structural connective corresponding to the quantifier, with the advantage that due to the presence of different structural connectives for the quantifiers (i.e., the nesting operator) and the propositional connectives (i.e., the comma and the sequent arrow) it is easier to adjust the system to other logics by varying the rules for the (structural) quantifiers. In particular, in *op. cit.* this gives rise to a nested sequent system for *free logic*, i.e., classical first-order logic where the domain might be empty.

### *11.4.2 Intuitionistic and Constructive Modal Logics*

Considering modal logics based on intuitionistic instead of classical propositional logic leads to the area of *intuitionistic and constructive modal logics*. Semantically, these logics are typically given by Kripke-frames with two accessibility relations $\leq$ and $R$, with the intuitionistic implication interpreted using the accessibility relation $\leq$ as usual, and the box modality understood as truth in all $R$-successors of all $\leq$-successors of the current world. The presence of two accessibility relations

then gives rise to a number of different possible interaction principles and in turn to a number of different possible logics (see, e.g., Simpson, 1994 for an introduction). Unfortunately, from the perspective of nested sequents, many of these interaction principles are similar to "diamond" or "confluence" principles such as $\forall x, y, z\,(xRy \wedge xRz \rightarrow \exists w.(yRw \wedge zRw))$ which gave rise to the extension of nested sequents to indexed nested sequents (see Sect. 11.5.6 below). Hence it is not clear whether they could be captured by combining the systems for modal logic from Sect. 11.2 with those for intuitionistic logic from Sect. 11.4.1, e.g., by considering two nesting operators for the modal and intuitionistic accessibility relations respectively.

Using a different approach, however, nested sequent systems for intuitionistic and constructive modal logics were proposed successfully by considering essentially a *single-conclusion* version of nested sequents. While the literature on nested sequents for these logics mainly makes use of the single-sided formulation of nested sequents, extended with a marker for a single designated *output formula*, i.e., the formula on the right hand side of the corresponding two-sided nested sequent, for the sake of consistency of notation here we stick to the two-sided notation. The system from Straßburger (2013) then reads as follows.

**Definition 11.4.1** An *intuitionistic nested sequent* is a nested sequent where the right hand side of exactly one component is non-empty and contains exactly one formula, called its *output formula*. The *output pruning* of an intuitionistic nested sequent $G$ is obtained by removing its (unique) output formula and is written as $G^{\downarrow}$.

It is worth noting that the modalities $\Box$ and $\Diamond$ are not interdefinable in intuitionistic and constructive modal logics. This together with the fact that an intuitionistic nested sequent contains only one output formula, and hence only one formula should appear on the right hand side of its formula interpretation leads to a slightly more involved definition of the latter:

**Definition 11.4.2** The *formula interpretation* of an intuitionistic nested sequent $G$ is written as $G^{\iota}$ and defined as follows, where $G_1, \ldots, G_n$ do not contain the output formula but $O$ does:

$$(M \Rightarrow [G_1], \ldots, [G_n])^{\iota} = \bigwedge M \wedge \Diamond(G_1^{\iota}) \wedge \cdots \wedge \Diamond(G_n^{\iota})$$
$$(M \Rightarrow [G_1], \ldots, [G_n], \phi)^{\iota} = \bigwedge M \wedge \Diamond(G_1^{\iota}) \wedge \cdots \wedge \Diamond(G_n^{\iota}) \rightarrow \phi$$
$$(M \Rightarrow [G_1], \ldots, [G_n], [O])^{\iota} = \bigwedge M \wedge \Diamond(G_1^{\iota}) \wedge \cdots \wedge \Diamond(G_n^{\iota}) \rightarrow \Box(O^{\iota})\,.$$

Thus the nesting operator is interpreted as the modal box on the right hand side of the implication whenever the structure inside contains the output formula, and as the modal diamond on the left hand side of the implication otherwise. The rules for the modal box then are essentially the same as in the classical setting:

$$\frac{G\{M, \Box\phi \Rightarrow N, [P, \phi \Rightarrow Q]\}}{G\{M, \Box\phi \Rightarrow N, [P \Rightarrow Q]\}}\,\Box_L \qquad \frac{G\{M \Rightarrow N, [\,\Rightarrow \phi]\}}{G\{M \Rightarrow N, \Box\phi\}}\,\Box_R$$

Since the modal operators are not inter-definable, in addition we also need the following rules for the modal diamond:

$$\frac{G\{M \Rightarrow N, [\phi \Rightarrow ]\}}{G\{M, \Diamond\phi \Rightarrow N\}}\ \Diamond_L \qquad \frac{G\{M \Rightarrow N, [P \Rightarrow Q, \phi]\}}{G\{M \Rightarrow N, [P \Rightarrow Q], \Diamond\phi\}}\ \Diamond_R$$

The intuitionistic character of the system is then captured on the level of propositional logic by adjusting the implication rules in the same way as is done for single-conclusion standard sequent systems for intuitionistic logic:

$$\frac{G^{\downarrow}\{M, \phi \to \epsilon \Rightarrow N, \phi\} \quad G\{M, \epsilon \Rightarrow N\}}{G\{M, \phi \to \epsilon \Rightarrow N\}}\ \to_L \qquad \frac{G\{M, \phi \Rightarrow N, \epsilon\}}{G\{M \Rightarrow N, \phi \to \epsilon\}}\ \to_R$$

Notice that the left premiss of the $\to_L$ rule contains $\phi$ as the output formula, so the output formula of $G$ needs to be removed by considering the output pruning $G^{\downarrow}$. This in turn may drastically change the formula interpretation of the subtree which contained the output formula before the output pruning. The remaining propositional rules are the standard ones. The resulting nested sequent system is shown to be sound and complete for intuitionistic modal logic IK in Straßburger (2013) via a cut elimination argument. The same work further contains extensions to all other logics of the intuitionistic version of the modal cube. Fully modular versions of these calculi are introduced in Marin and Straßburger (2014), and the systems are adapted to constructive modal logics in Arisaka et al. (2015). An implementation is presented in Girlando and Straßburger (2020). A Maehara-style multi-conclusion version of the calculi was introduced in Kuznets and Straßburger (2018) and used to obtain countermodels from failed proof search. Calculi for the more general class of logics extending IK with seriality and so-called *Horn Scott–Lemmon axioms*, i.e., axioms of the shape $(\Diamond^n\Box A \to \Box^k A) \wedge (\Diamond^k A \to \Box^n\Diamond A)$ were proposed recently in Lyon (2021a). The single-conclusion nested sequent calculi for the logics of the intuitionistic modal cube were also developed independently in Galmiche and Salhi (2018) under the name of *tree sequents*, based on the natural deduction systems of Galmiche and Salhi (2010).

### *11.4.3 Other Non-standard Normal Modal Logics*

Modal logic was firstly introduced to formalize the concepts of necessity and possibility; however, soon after the introduction of Kripke possible worlds semantics modal logic was used to formalize other notions such as knowledge, belief or obligation. In accordance with these new interpretations of the modal operator, new logical systems have been developed. We use this subsection to briefly present the application of nested calculi to some of these alternative interpretations of $\Box$ and $\Diamond$.

An important interpretation of the modalities is given in the context of *modal tense logics*, where $\Box\phi$ is interpreted as "at every point in the future it will be the

case that $\phi$". Dually, $\Diamond\phi$ is interpreted as "at some point in the future it will be the case that $\phi$". Analogously to these *forward looking* modalities, modal tense logics also contain the *backward looking* modality $\blacksquare$ and its dual $\blacklozenge$, interpreted as "at every point in the past it was the case that" and "at some point in the past it was the case that", respectively. As usual, $\Diamond$ and $\blacklozenge$ can be taken to be defined in terms of $\Box$ and $\blacksquare$, respectively. The tense analogue of modal logic $\mathsf{K}$ is denoted $\mathsf{K}_t$ and given by Hilbert-style axiomatisations of modal logic $\mathsf{K}$ for both the forward and the backward looking modalities together with the axioms $\phi \to \Box\blacklozenge\phi$ and $\phi \to \blacksquare\Diamond\phi$ for the connection between the forward and backward looking modalities.

Historically, modal tense logics were one of the areas which sparked the creation of the nested sequent framework in Kashima (1994) (originally formulated in the one-sided setting, see Sect. 11.2.1). For capturing both the forward and backward looking modalities, nested sequents for modal tense logics contain a *forward* or *future* nesting operator ${}^F[.]$ as well as a *backward* or *past* nesting operator ${}^P[.]$. The rules for the basic logic $\mathsf{K}_t$ are essentially the nested sequent rules for $\mathsf{K}$ for both modalities $\Box$ and $\blacksquare$ together with the following two rules for their interaction:

$$\frac{G\{\Gamma, \phi \Rightarrow \Delta,\ {}^F[\Sigma \Rightarrow \Pi]\}}{G\{\Gamma \Rightarrow \Delta,\ {}^F[\Sigma, \blacksquare\phi \Rightarrow \Pi]\}} \qquad \frac{G\{\Gamma, \phi \Rightarrow \Delta,\ {}^P[\Sigma \Rightarrow \Pi]\}}{G\{\Gamma \Rightarrow \Delta,\ {}^P[\Sigma, \Box\phi \Rightarrow \Pi]\}}$$

Nested sequent calculi for basic extensions of modal tense logic with frame properties like reflexivity, transitivity, connectedness or totality were also already considered in Kashima (1994). However, very general completeness results for large classes of modal tense logics were given later. In particular, Goré et al. (2011) obtain cut-free deep inference calculi for modal tense logics given by *Scott–Lemmon axioms*, i.e., axioms of the form $\Diamond^h\Box^i\phi \to \Box^j\Diamond^k\phi$ for $h, i, j, k \geq 0$ (see also Sect. 11.5.6 below), some of which also yield cut-free nested sequent calculi, and for modal tense logics given by *path axioms*, i.e., axioms of the form $\langle\rangle_1 \ldots, \langle\rangle_n\phi \to \langle\rangle\phi$, where $n \geq 0$ and each of $\langle\rangle_1, \ldots, \langle\rangle_n, \langle\rangle$ is either $\Diamond$ or $\blacklozenge$. Since normal modal logics can be understood as the restriction of modal tense logic to the forward looking modalities, by appropriate conservativity results shown in *op. cit.* the calculi for the Scott–Lemmon axioms also give rise to nested sequent calculi for the corresponding normal modal logics. It should be noted, however, that these calculi might still contain the backward looking structural connective, and hence their structures might not have a formula interpretation in the language of standard (non-tense) normal modal logic. Even further, general completeness results for *grammar logics*, i.e., multimodal tense logics, where loosely speaking the connections between the modalities are given by rewriting rules, were shown in Tiu et al. (2012). For space reasons we do not consider all the details here. The interested reader is referred to the mentioned literature.

The second non-alethic interpretation of modal logic that we present in this section is the epistemic interpretation, namely the interpretation of the $\Box$ in terms of "it is known that." More precisely, in order to express the fact that different agents may have different knowledges, we add indexes to the $\Box$: one index for each agent of the group of agents considered in the language. The system that seems to best represent

an epistemic multiagent interpretation of modal logic is S5. We can implement the multi-agent version of S5 by adding a new operator $\circledast$ for common knowledge, where common knowledge is standardly defined as the infinite conjunction "all agents know $\phi$, and all agents know that all agents know $\phi$ and so on." We will denote the multi-agent version of the Hilbert system S5 plus common knowledge by the name CS5. In CS5 we capture common knowledge by means of a fixed point axiom, which states that common knowledge is a fixed point, and an induction rule that states that this fixed point is the greatest fixed point. From a semantic point of view, the common knowledge operator is formally defined as the modality of reachability that uses accessibility edges corresponding to any of the knowledge operators for the agents.[10]

In order to get a nested calculus for CS5, the first mandatory step is to index nested sequents with agents. However, as explained in Poggiolesi (2010c), there are two nested calculi for S5: one that simply uses hypersequents (if we look at hypersequents as nested sequents where the accessibility relation has collapsed, see Sect. 11.5.2.) and one with full nested sequents apparatus that include rules for reflexivity, transitivity, and asymmetry (see Sect. 11.2). Since the former is simpler, it seems the more adequate to work with in the epistemic context. Hence, first of all, the hypersequents for S5 are indexed with indexes denoting agents,[11] and then rules for the operator of common knowledge are introduced.

The difficulty of dealing with the common knowledge operator is that it has deep infinitary features; hence at the Gentzen level, it has been taken into account in either of the following two ways: either with finitary calculi (that often do not enjoy cut-elimination, e.g., see Alberucci & Jager, 2005) , or with infinitary rules (e.g., see Tanaka, 2003). In the nested framework both ways have been explored. We will briefly present them.

**Definition 11.4.3** We consider a language $\mathcal{L}_h^{\Box}$ with a finite set $\Phi$ of agents {a, b, c, …}. Propositions $S$ are atoms. Formulae are given by the following grammar:

$$\phi ::= S \mid \neg\phi \mid (\phi \wedge \phi) \mid \Box_z\phi \mid \circledast\phi$$

where $z \in \Phi$, the formula $\Box_z\phi$ is read as "agent $z$ knows $\phi$" and the formula $\circledast\phi$ is read as "$\phi$ is common knowledge". We will use the formula $\Box\phi$ as an abbreviation for "everybody knows $\phi$":

$$\Box\phi = \Box_a\phi \wedge \cdots \wedge \Box_z\phi$$

In Brünnler and Stüder (2009) the following two rules for common knowledge in the framework of nested calculi are considered[12]:

[10] For the Hilbert-system and the semantics for CS5, e.g., see Fagin et al. (2004).

[11] See also Sect. 11.5.6.

[12] In the rules, the notation $A$ stands for a set of agents that satisfy certain features. We omit to describe these features here for the sake of brevity; they can be found in Hill and Poggiolesi (2015).

$$\frac{G\{A : \Box^k\phi, M \Rightarrow N\}}{G\{A : \circledast\phi, M \Rightarrow N\}}\ \circledast L \qquad \frac{G\{A : M \Rightarrow N, \Box^k\phi\}\ for\ all\ k \geqslant 1}{G\{A : M \Rightarrow N, \circledast\phi\}}\ \circledast R$$

Although these rules allow a cut-elimination procedure, the rule $\circledast R$, being infinitary, vanish any proof-search attempt. A more fine-grained result can then be found in Hill and Poggiolesi (2015), where common knowledge is treated by means of the following three rules in an hypersequent framework:

$$\frac{G \mid A : \circledast\phi, \Box\phi, M \Rightarrow N}{G \mid A : \circledast\phi, M \Rightarrow N}\ \circledast L_1$$

$$\frac{G \mid A : \circledast\phi, M \Rightarrow N \mid B : \circledast\phi, P \Rightarrow Q}{G \mid A : \circledast\phi, M \Rightarrow N \mid B : P \Rightarrow Q}\ \circledast L_2$$

$$\frac{G \mid A : M \Rightarrow N, \epsilon \quad \epsilon \Rightarrow \Box\phi \quad \epsilon \Rightarrow \Box\epsilon}{G \mid A : M \Rightarrow N, \circledast\phi}\ \circledast R$$

The rules are finitary, the calculus they belong to is cut-free; moreover, in order to control proof-search, it has been shown that the rule $\circledast R$ can be restricted to a certain type of formulae only, i.e., it has been shown that the formula $\epsilon$ occurring in the premiss can only be of a certain form. Details can be found in Hill and Poggiolesi (2015).

Let us now move to a third different interpretation of the modal operator and related logic, incidentally one of the first treated by the nested sequent method in Bull (1992). Propositional dynamic logic, or PDL for short, is a (modal) logic based on the idea of associating with each program term $a$ of a programming language a modality $[a]$. This means that in PDL we still deal with boxed formulae as we do in modal logic, but the box is no longer empty but filled with program terms.

Several different programs can fill the box: atomic programs ($a_0$, $a_1$, $a_2$, ...), but also more complex ones, which are denoted by the capital letters $A, B, C \ldots$. These latter are constructed by means of the following program operators: the union operator, $A \cup B$, that should be interpreted as: "do $A$ or $B$ non-deterministically;" the composition operator, $A \otimes B$, that should be interpreted as: "first do $A$ and then do $B$;" the test operator, $\phi?$, that should be interpreted as: "verify that $\phi$ is true;" and finally the iteration operator, $A^*$, that should be interpreted as: "repeat $A$ a finite number of times." In PDL, we thus deal with formulae of the following form: $[a]\phi$, $[A \cup B]\phi$, $[A \otimes B]\phi$, $[\epsilon\ ?]\phi$, $[A^*]\phi$, each of which should be read as: "$\phi$ is true after every terminating execution of the program that is in the box."

From the point of view of Hilbert systems, propositional dynamic logic is well-defined. Indeed, the system for PDL is obtained by adding to classical propositional logic: (i) the distribution axiom schema, that has the form: $[A](\phi \rightarrow \epsilon) \rightarrow ([A]\phi \rightarrow [A]\epsilon)$, for each program $\phi$; (ii) modus ponens and the rule of necessitation; and (iii) at least one axiom schema or inference rule for each program operator. As for the Gentzen perspective, the steps and difficulties to formulate a satisfactory nested

Note also that in Brünnler and Stüder (2009) the rules for common knowledge are added to the multiagent version of the nested calculus for $\mathbf{NS_k}$. However, for the sake of uniformity, they can also be thought as added to the multiagent version of the hypersequent system for S5.

calculus are analogous to those for common knowledge.[13] On the one hand, just as in the epistemic case nested sequents were indexed by agents, now nested sequents are indexed by (atomic) programs; on the other hand, just as the common knowledge operator involved an infinitary rule, so it is the case in PDL for the program operator $*$. Hence the nested calculus for PDL is obtained by adding to the usual initial nested sequents and nested rules for the propositional and modal connectives,[14] the following rules[15]:

$$\frac{G\{[B]\phi, [C]\phi, M \Rightarrow N\}}{G\{[B \cup C]\phi, M \Rightarrow N\}} \cup L \qquad \frac{G\{M \Rightarrow N, [B]\phi\} \quad G\{M \Rightarrow N, [C]\phi\}}{G\{M \Rightarrow N, [B \cup C]\phi\}} \cup R$$

$$\frac{G\{[B][C]\phi, M \Rightarrow N\}}{G\{[B \otimes C]\phi, M \Rightarrow N\}} \otimes L \qquad \frac{G\{M \Rightarrow N, [B][C]\phi\}}{G\{M \Rightarrow N, [B \otimes C]\phi\}} \otimes R$$

$$\frac{G\{M \Rightarrow N, \phi\} \quad G\{\epsilon, M \Rightarrow N\}}{G\{[\phi?]\epsilon, M \Rightarrow N\}} ?L \qquad \frac{G\{\phi, M \Rightarrow N, \epsilon\}}{G\{M \Rightarrow N, [\phi?]\epsilon\}} ?R$$

$$\frac{G\{[B^*]\phi, [B]^n\phi, M \Rightarrow N\}}{G\{[B^*]\phi, M \Rightarrow N\}} *L \qquad \frac{G\{M \Rightarrow N, [B]^n\phi\} \ \textit{for each } n < \omega}{G\{M \Rightarrow N, [B^*]\phi\}} *R$$

In the calculus so obtained all structural rules are (height-preserving) admissible, the logical rules are invertible and the calculus is cut-free, see Hill and Poggiolesi (2010).

Let us end the section by briefly mentioning some further applications of the nested sequent method. Nested sequent calculi for a class of multimodal logics called *simply dependent multimodal logics* can be obtained from the linear nested sequent calculi (see also Sect. 11.5.3) for these logics constructed in Lellmann and Pimentel (2019). However, since these logics all have standard sequent calculi already, we do not consider the details here but refer the interested reader to *op. cit.* instead. Another application concerns the modal logic of provability—also called the logic GL from the initials of Gödel and Löb—which is the modal logic resulting from the interpretation of the $\Box$ in terms of "it is provable in Peano Arithmetic that." The Hilbert system GL is obtained by adding to the system K4 the GL axiom: $\Box(\Box\phi \to \phi) \to \Box\phi$. The nested sequent method can be successfully applied to the system GL. Since this logic provides a good showcase of the connection between

[13] Note that if PDL is considered without the program operator $*$, then the first nested sequent calculus for it has been formulated by Bull (1992).

[14] Modal rules are adapted to the fact that the $\Box$ contains a program.

[15] Note that in both the rules $*L$ and $*R$, we use the notation $[B]^n\phi$, which is inductively defined in the following way:

- $[A]^0\phi := \phi$
- $[A]^{k+1}\phi := [A][A]^k\phi$.

the standard and nested sequent frameworks we will discuss the resulting system in Sect. 11.5.1 below. A system for the logic GLP, a different flavor of provability logic related to ordinal analysis of arithmetic, is developed in Shamkanov (2015). Nested sequents have also been applied to justification logic, the logic introduced by Artemov (2008), where the $\Box$ is substituted by terms or justifications that have the the form $t, f, r, \ldots$. Whilst in Fitting (2012) one may find a nested calculus for justification logic which is cut-free but not analytic, in Hill and Poggiolesi (2019) hypersequents are used to provide a cut-free and analytic calculus for justification logic. However, the process to obtain analyticity is quite long and laborious so we completely omit it here.

### *11.4.4 Non-normal Modal Logics and Conditional Logics*

All the modal logics considered so far share a common property, whether based on classical or intuitionistic propositional logic: they are *normal* in the sense that they satisfy the necessitation rule, or alternatively, that the formula $\Box\top$ is a theorem. However, already in Kripke (1965), the Kripke semantics was extended to cover also *non-normal* modal logics. In fact, historically many of the first systems proposed for modal logic were non-normal (see the references in *op. cit.*). Thus it is an interesting question whether the nested sequent formalism can be extended to cover non-normal modal logics as well. Unfortunately, for non-normal modal logics the parallel between the tree structure of the models and the tree structure of nested sequents breaks down to some extent, since non-normality of the modal logic necessitates additional structure on the models. This could be the addition of *non-normal worlds*, where the formula $\Box\top$ is false in Kripke (1965), or the more fine-grained analysis of the accessibility relation in what is known under the alternative names of *neighbourhood semantics*, *Scott–Montague semantics* or *minimal model semantics*, see, e.g., Chellas (1980). Perhaps even more problematic is the fact that in every nested sequent system which contains the standard (local) propositional rules of Sect. 11.2 and interprets the nesting operator [.] as a connective $\circ$ using rules analogous to the standard left and right rules for $\Box$ we have the following derivations:

$$\dfrac{}{\dfrac{\Rightarrow [\Rightarrow \top]}{\Rightarrow \circ\top}\,\circ_R} \qquad \dfrac{\dfrac{\dfrac{}{\circ p, \circ q \Rightarrow [p, q \Rightarrow p]} \quad \dfrac{}{\circ p, \circ q \Rightarrow [p, q \Rightarrow q]}\,\wedge_R}{\circ p, \circ q \Rightarrow [p, q \Rightarrow p \wedge q]}}{\dfrac{\circ p, \circ q \Rightarrow [\Rightarrow p \wedge q]}{\circ p, \circ q \Rightarrow \circ(p \wedge q)}\,\circ_R}\,\circ_L, \circ_L$$

Hence in every such system we can derive the axioms $\circ\top$ as well as $\circ p \wedge \circ q \rightarrow \circ(p \wedge q)$. Similarly, every such system also derives the converse of the latter, i.e., $\circ(p \wedge q) \rightarrow \circ p$. These three axioms together, however, characterise $\circ$ as a normal modality. Hence it is not possible to simply reinterpret the nesting operator as a non-normal modality while keeping the standard propositional rules intact.

While the development of a general proof theory for non-normal modal logics in the nested sequent framework and in particular a general link to the semantics is still an open problem, first steps were taken in Lellmann (2019). The main idea in that approach is to make use of the fact from Brown (1988) that neighbourhood structures can be seen to give rise to both a non-normal and a normal modality at the same time. On the level of nested sequents this means that in order to capture certain non-normal modal logics it is enough to add a second nesting operator $\langle . \rangle$ which acts as a structural connective for the non-normal modality. To model both the normal and the non-normal modality arising from neighbourhood structures, we extend the language to include the (non-normal) modality $\langle ]$ as well as the normal modality $\Box$. Note that due to the inherently normal nature of any interpretation of the standard nesting operator [.] as considered above we can only obtain a formula interpretation of nested sequents with this operator if we include a normal modality in the language, or if in the considered logic it is definable in terms of the non-normal modality. The latter is in general not the case.

**Definition 11.4.4** A *non-normal nested sequent* is a finite object

$$M \Rightarrow N, \langle P_1 \Rightarrow Q_1 \rangle, \ldots, \langle P_n \Rightarrow Q_n \rangle, [G_1], \ldots, [G_m]$$

where $M \Rightarrow N$ as well as $P_i \Rightarrow Q_i$ for $i \leq n$ is a sequent and where $G_j$ for $j \leq m$ is a non-normal nested sequent. Its formula interpretation $G^\Delta$ is given by:

$$G^\Delta := \bigwedge M \to \bigvee N \vee \bigvee_{i \leq n} \langle ](\bigwedge P_i \to \bigvee Q_i) \vee \bigvee_{j \leq m} \Box (G_j)^\Delta .$$

The crucial point is that in a non-normal nested sequent other nested sequents can occur only in the scope of the normal nesting operator [.], not the non-normal one $\langle . \rangle$. Further, in order to prevent the derivation of the axioms for normality above, the application of all rules needs to be restricted not to occur in the scope of the non-normal nesting operator $\langle . \rangle$. The resulting modal rules then are the standard rules $\Box_L$ and $\Box_R$ for the normal operator together with the left and right rules for the non-normal modality

$$\frac{G\{M \Rightarrow N, [P, \phi \Rightarrow Q]\}}{G\{M, \langle ]\phi \Rightarrow N, \langle P \Rightarrow Q \rangle\}} \langle ]_L \qquad \frac{G\{M \Rightarrow N, \langle \Rightarrow \phi \rangle\}}{G\{M \Rightarrow N, \langle ]\phi\}} \langle ]_R$$

and a single *interaction rule* for the relation between the normal and non-normal modalities:

$$\frac{G\{M \Rightarrow N, [P \Rightarrow Q]\}}{G\{M \Rightarrow N, \Box\phi, \langle P \Rightarrow Q \rangle\}} \; I \; .$$

This rule corresponds to the interaction axiom $\Box\epsilon \to \Box\phi \vee \langle ]\epsilon$ and essentially states (seen bottom-up) that whenever there is any boxed formula on the right hand side of the conclusion, then a non-normal nesting operator can be converted into a normal one.

Note that since no rule can be applied inside $\langle . \rangle$, and the only rules which mention this operator in the conclusion change it to the normal nesting operator in the premiss, essentially the only property which can be derived for the language with only the non-normal modality $\langle ]$ is $\langle ]\phi \to \langle ]\epsilon$ for derivable $\phi \to \epsilon$. Correspondingly, this system restricted to this language yields a nested sequent system for non-normal *monotone modal logic*. The system was further modified to cover some extensions in *op. cit.* Purely syntactic systems exploiting the idea of translating standard sequent calculi to a (linear) nested framework (see Sect. 11.5.1 below) were also considered in Lellmann and Pimentel (2019). This approach yields modular systems for a non-normal analogue of the modal cube, the *modal tesseract*, but has the disadvantage that in absence of a second normal modality the (linear) nested sequent structures do not seem to have a formula translation.

Finally, a different flavour of non-normal modal logics is given by *conditional logics*, or logics of *counterfactual implication*. The weakest of these logics extend classical propositional logic with a binary modality $(\phi > \epsilon)$ for *conditional implication*, read as "if $\phi$ were the case, then also $\epsilon$ would be the case". This modality behaves like normal modal logic $\mathsf{K}$ in the second argument, but only contains the *congruence rule* for the first argument: if $\phi \leftrightarrow \epsilon$ is derivable, then so is $(\phi > \delta) \to (\epsilon > \delta)$. Nested sequents for the basic conditional logic $\mathsf{CK}$ and a number of extensions were introduced in Alenda et al. (2013), making use of a dyadic nesting operator $[. : .]$, where the first argument contains a single formula, and the second argument contains a nested sequent.

**Definition 11.4.5** A *conditional nested sequent* is a finite structure

$$M \Rightarrow N, [\phi_1 : G_1], \ldots, [\phi_n : G_n]$$

where $M \Rightarrow N$ is a sequent, the $\phi_i$ are formulae and the $G_i$ are nested sequents. Its formula interpretation $G^\Delta$ is $\bigwedge M \to \bigvee N \vee (\phi_1 > (G_1)^\Delta) \vee \cdots \vee (\phi_n > (G_n)^\Delta)$.

Thus, again only the "normal" part of the nesting operator can contain further nesting operators. The rules for the propositional connectives then are standard (restricted to the second argument of the nesting operators). The basic modal rules, reformulated from the original one-sided framework to the two-sided framework of this paper, are given by:

$$\frac{G\{M \Rightarrow N, [\phi : P, \epsilon \Rightarrow Q]\} \quad \phi \Rightarrow \delta \quad \delta \Rightarrow \phi}{G\{M, \phi > \epsilon \Rightarrow N, [\delta : P \Rightarrow Q]\}} >_L \qquad \frac{G\{M \Rightarrow N, [\phi : \Rightarrow \epsilon]\}}{G\{M \Rightarrow N, \phi > \epsilon\}} >_R$$

For the sake of clarity we omitted copies of the principal formulae in the premisses. This system along with a number of extensions was used in Alenda et al. (2013) in a syntactical cut elimination procedure and yielded complexity results via backward proof search.

A nested sequent system for a much stronger conditional logic, rephrased as a logic of conditional belief, is given in Girlando et al. (2019). Since in this logic a knowledge operator is definable, it contains a multi-agent version of modal logic $\mathsf{S5}$.

On the nested sequent level this is mirrored by the fact that the system essentially contains a multi-agent version of the nested sequent calculus with the rules $t$, 4 and 5 from Sect. 11.2.

## 11.5 Relation to Other Formalisms

In order to get a clearer understanding of the benefits and shortcomings of the nested sequent framework in general, it is helpful to situate it with respect to other important proof-theoretic frameworks. In the following we consider the relations to those well-known frameworks which to the best of our knowledge are most often used in the context of modal logics. In particular, these are the frameworks of ordinary sequents (Sect. 11.5.1), hypersequents (Sect. 11.5.2), prefixed tableaux (Sect. 11.5.4), labelled sequents (Sect. 11.5.5) and display calculi (Sect. 11.5.7). In addition we also consider two modifications of the nested sequent framework in the form of linear nested sequents and indexed nested sequents (Sects. 11.5.3 and 11.5.6, respectively). These frameworks are interesting in that they provide a middle-ground in expressivity between standard and nested sequents (linear nested sequents), and extend the expressivity of nested sequents toward that of labelled sequents (indexed nested sequents). The relative size of the different sections of course by no means reflects the relative volume of work in the corresponding areas or their importance. The relations between the frameworks are summarized in Fig. 11.9. For space reasons unfortunately we have to omit the interesting connections to frameworks like *multiple sequent calculi* (e.g., Indrzejczak, 1997), *higher-arity sequent calculi* (e.g., Blamey & Humberstone, 1991), *modal tree-sequents* (e.g., Cerrato, 1993) and *indexed sequent calculi* (e.g., Mints, 1997). The reader is referred to Poggiolesi (2010b) for more details. To avoid unnecessary and potentially confusing notation in this section we omit names for rules where they are only mentioned or used locally. For space reasons we also do

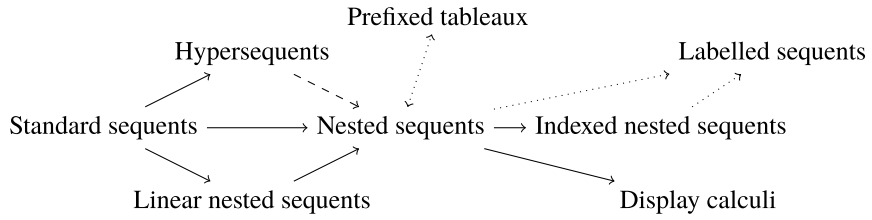



**Fig. 11.9** The proof-theoretic frameworks considered in this section with their relations. Solid arrows indicate that the structures manipulated in one framework can also be found in the other framework. Dotted arrows indicate that the structures of one framework can be translated into structures in the other framework. The dashed arrow indicates that there is an interpretation (see Sect. 11.5.2 for details)

not give all the details on soundness and completeness results about specific logics in the different proof-theoretic frameworks in the form of theorems, but refer to the literature instead.

### *11.5.1 Sequents*

As mentioned above, the nested sequent framework is a direct extension of the original *sequent* framework, introduced in Gentzen (1935). In particular, standard sequents can be seen as nested sequents without occurrences of the nesting operator [.]. Because of this, every standard sequent calculus can also be seen as a nested sequent calculus. However, in such calculi none of the additional structures and features of nested sequents are used. Hence doing so means relinquishing the benefits of the nested sequent framework, in particular modularity, and so does not gain much over staying in the original sequent framework.

A deeper connection stems from the the fact that the modal rules in the standard sequent framework can be seen as *transitional rules* (see, e.g., Goré, 1999), which in the context of countermodel construction from a failed proof search correspond to moving to a successor world. Since a successor world in the Kripke semantics corresponds directly to a successor node in a nested sequent, this suggests a straightforward simulation of the modal or transitional rules for standard sequents in terms of the modal rules for nested sequents as shown in the following example.

**Example 11.5.1** The application of the modal rule in a standard sequent system for modal logic $\mathsf{K}$ (below on the left) is simulated by the sequence of rule applications in the nested sequent system $\mathbf{NS}^{+}_{\mathsf{k}}$ for $\mathsf{K}$ (below on the right).

$$\frac{\phi_1, \ldots, \phi_n \Rightarrow \epsilon}{\Box\phi_1, \ldots, \Box\phi_n \Rightarrow \Box\epsilon}\, k \quad \rightsquigarrow \quad \dfrac{\dfrac{\Box\phi_1, \ldots, \Box\phi_n \Rightarrow \Box\epsilon\,, [\phi_1, \ldots, \phi_n \Rightarrow \epsilon\,]}{\Box\phi_1, \ldots, \Box\phi_n \Rightarrow \Box\epsilon\,, [\,\Rightarrow \epsilon\,]}\, \Box R^{+}}{\Box\phi_1, \ldots, \Box\phi_n \Rightarrow \Box\epsilon}\, \Box L^{+}$$

Here the double lines in the derivation on the right stand for multiple applications of the rule $\Box L$. Note that in the premiss of the derivation on the right the root sequent contains exactly the conclusion of the application of the modal rule on the left, while the successor node in this nested sequent contains exactly the premiss of the application of the sequent rule.

The same principle can be applied to all the logics of the modal cube (Fig. 11.2) which have a cut-free standard sequent system, i.e., the logics $\mathsf{K}$, $\mathsf{KD}$, $\mathsf{KT}$, $\mathsf{K4}$, $\mathsf{KD4}$, $\mathsf{S4}$, $\mathsf{K45}$, $\mathsf{KD45}$. Given that the latter calculi are complete for the respective logics, this yields straightforward alternative completeness proofs for these systems:

**Proposition 11.5.2** *For any $X \subseteq \{d, t, 4\}$ the nested sequent system $\mathbf{NS}^{+}_{\mathsf{k}} \cup X$ is cut-free complete for the corresponding modal logic. The same holds for $X = \{4, 5\}$ and $X = \{d, 4, 5\}$.*

***Proof*** By simulating the sequent rules at the appropriate node in the nested sequent. □

Note that this approach can also be used to convert standard sequent calculi into nested sequent calculi: Identify the part of the standard sequent rule which is responsible for the noninvertible behaviour (the decision on the boxed formula on the right in the rule $k$ above) and turn this into a successor-creating rule. Then, build up the rest of the premiss(es) by unpacking/transferring one formula at a time into the successor node (the boxed formulae on the left in the rule $k$ above). See also, e.g., Lellmann (2015) and Lellmann and Pimentel (2019).

The idea of simulating sequent derivations using the nested sequent seems to have been mentioned first in Brünnler (2009) for modal logic KT. It was systematically explored in Lellmann (2015), Lellmann and Pimentel (2019) and used also in Parisi (2017) and Parisi (2020). While the latter references are in the intermediate framework of *linear nested sequents* (see Sect. 11.5.3), the results apply to the full nested sequent framework as well. A similar approach is also taken in Marin et al. (2016), where derivations in a number of different frameworks are simulated in a focused labelled sequent system (see also Sect. 11.5.5). In the form presented here, the idea of simulating sequent derivations by nested sequent derivations leads to a number of observations, listed in Lellmann (2015) and Lellmann and Pimentel (2019).

First, while, e.g., in the standard sequent system for modal logic K4, the modal rule is changed to accommodate the boxed formulae in the premiss, in the nested sequent system it is enough to just add a single rule which can be used to copy the boxed formulae into the successor node one by one. This is due to the fact that the standard sequent rules are essentially broken down into smaller parts in the nested sequent framework. Hence nested sequent systems for these logics can be seen as a *modularization* of standard sequent systems.

Second, since the simulation of a standard sequent derivation does not introduce any branching in the resulting nested sequents, for completeness with respect to these simpler rules it is enough to consider non-branching nested sequents. This idea is further explored in the framework of *linear nested sequents*, see also Sect. 11.5.3.

Third, since the simulation of standard sequent rules always works on a leaf node of the corresponding nested sequent, the corresponding nested sequent calculi can be restricted to their *end-active* version, where in every rule application one of the active nodes of the nested sequent in the premiss(es) must be a leaf node.

Fourth, given that a node and its successor node in a nested sequent resulting from translating a sequent rule can be seen as the conclusion and premiss of that rule, the branches of such a nested sequent can be viewed essentially as *traces* or *histories* of depth-first backward proof search in the corresponding sequent system. Here, the conclusion of a noninvertible (i.e., modal or transitional) standard sequent rule is stored in the nested sequent before moving on to the premiss, i.e., the successor node in the nested sequent. This in turn suggests that creating branching in a nested sequent by applying a successor-creating rule such as $\Box R$ to a non-leaf node of the nested sequent loosely corresponds to the creation of a *backtracking point* in the backward proof search for the standard sequent system: One of the noninvertible

rules is applied (backward), and its premiss can be further expanded. Backtracking to this backtracking point in the proof search and applying a different noninvertible rule then roughly corresponds to applying a different successor-creating rule to the node. In particular, the branching in the nested sequent then corresponds to *OR-branching* in the proof search tree. An advantage of the nested sequent formulation is however, that the current branch is not discarded when applying such a rule, as would be the case for backtracking in depth-first backward proof search. In fact, the procedure $prove(G, \mathbf{NS}^{+}_{\mathsf{K}})$ of Sect. 11.3 for proof search in nested sequents, which first applies all rules which do not create successor nodes, and then creates an additional layer by applying $(\Box R)^*$ wherever possible once, corresponds to *breadth-first* proof search in the corresponding standard sequent system: First, all invertible rules are applied to the sequent, then every possible noninvertible rule is applied. The latter applications yield a set of sequents corresponding essentially to the leaves of the nested sequent obtained by applying the rule $(\Box R)^*$ wherever possible. Then the process is repeated on this set of sequents, i.e., the new leaf nodes of the nested sequent. This suggests that the nested sequent framework can also be used as a tool to analyse different forms of proof search in the standard sequent framework, an area of research which apart from the basic notions has not yet been fully explored.

The previous observations are based on enhancing or studying existing standard sequent calculi, using translations and simulations in calculi in the nested sequent framework. Moreover, the fact that the nested sequent framework can capture logics for which no cut-free sequent calculus is known, such as modal logics S5 or K5, suggests that the nested sequent framework is stronger than the standard sequent framework. In light of this it is also interesting to investigate the fragment of nested sequent calculi corresponding to standard sequent calculi, and in particular which kinds of nested sequent calculi can be converted back into standard sequent calculi.

In this spirit, some sufficient criteria for converting a nested sequent calculus back into a standard sequent calculus were introduced in Pimentel et al. (2019). The main idea again stems from the connection between the nesting operator in the nested sequent framework and the transitional rules in the standard sequent framework, and is based on reordering nested sequent derivations by rule permutations so that the modal rules are grouped in *blocks* consisting (when seen bottom-up from conclusion to premisses) of a single application of a *creation rule*, i.e., a nested sequent rule introducing a new nesting operator, followed by a number of *upgrade rules*, i.e., rules which move (possibly modified) formulae into the newly created node. The properties of a nested sequent system which ensure that this can be done then are:

1. the creation rules introduce exactly one nesting operator;
2. the upgrade rules never move a formulae toward the root of the nested sequent;
3. the upgrade rules move a formula inside exactly one nesting operator.

Nested sequent calculi satisfying these criteria are called *basic nested sequent calculi*. These properties together entail that such a block consisting of a creation rule and a number of upgrade rules can be converted into a standard sequent rule. Moreover, the resulting rule follows the general format of *basic sequent calculi* of Lahav and Avron (2013) (see *op. cit.* for the definition). Hence, we have:

**Theorem 11.5.3** (Pimentel et al., 2019, Corollary 30) *For every basic nested sequent system, there is an equivalent standard sequent system.*

It is instructive to consider how, e.g., nested sequent calculi for modal logics like S5 and KB, which are notoriously difficult in the (cut-free) standard sequent setting, fail these criteria. E.g., the rules

$$\frac{G\{\Box\phi, M \Rightarrow N, [\Box\phi, P \Rightarrow Q]\}}{G\{M \Rightarrow N, [\Box\phi, P \Rightarrow Q]\}}\ 5 \quad \text{or} \quad \frac{G\{\phi, M \Rightarrow N, [\Box\phi, P \Rightarrow Q]\}}{G\{M \Rightarrow N, [\Box\phi, P \Rightarrow Q]\}}\ b$$

considered above move the formula $\Box\phi$ toward the root of the nested sequent and hence fail property 2 above.

As a more explicit example for the relation between standard and nested sequent systems, consider modal provability logic GL. As already mentioned in Sect. 11.4.3 above, this logic is given by adding the axiom $\Box(\Box\phi \to \phi) \to \Box\phi$ to normal modal logic K4. The (Kleene'd version of the) nested sequent calculus CSGL for provability logic GL from Poggiolesi (2009) is given by the modal rules in Fig. 11.10. Since the rules satisfy all the criteria for sequentialisation from Pimentel et al. (2019), we can rearrange derivations in the calculus CSGL into blocks consisting of an application of the creation rule $\Box K_{\mathsf{gl}}$ at the bottom to create a new component $[\Box\phi \Rightarrow \phi]$ in the nested sequent, followed by applications of the rules $\Box A_{\mathsf{gl}}$ and 4 to populate this component. The sequent rule corresponding to such a block is (using the set notation to denote the set of premisses given by sets $R_1$, $R_2$ with $R_1 \cup R_2 = R$ and $R$ as defined below):

$$\frac{\{\ R_1, \Box S, \Box\phi \Rightarrow \Box R_2, \phi : R_1 \cup R_2 = R\ \}}{\Box M, P \Rightarrow N, \Box\phi}$$

where $\Box R \subseteq \Box M$ is the multiset of formulae to which the nested sequent rule $\Box A_{\mathsf{gl}}$ was applied in the block, and $\Box S \subseteq \Box M$ is the multiset of formulae to which the nested sequent rule 4 was applied. The large number of premisses for this rule (given by all the possibilities for sets $R_1$, $R_2$ with $R_1 \cup R_2 = R$) stems from the fact that the nested sequent rule $\Box A_{\mathsf{gl}}$ has two premisses, and hence every application of that rule doubles the number of premisses of the corresponding sequent rule. Thus a proof of cut-free completeness for the nested sequent calculus for GL automatically yields cut-free completeness of the corresponding standard sequent calculus, i.e., the sequent calculus with the standard propositional rules and this rule. Note that considering $R_2 = \emptyset$ in the rule above, that rule contains in particular the premiss $R, \Box S, \Box\phi \Rightarrow \phi$. But then we can further replace any application of this rule in a derivation by an application of the simpler rule

$$\frac{\Box M, N, \Box\phi \Rightarrow \phi}{\Box M, \Box N \Rightarrow \Box\phi}\ GLR$$

(possibly modulo contraction) just by deleting the derivations of the remaining premisses. Since this is (a slight variant of) the standard modal sequent rule for GL (see,

$$\frac{G\{\Box\alpha, M \Rightarrow N, [P \Rightarrow Q, \Box\alpha]\} \quad G\{\Box\alpha, M \Rightarrow N, [\alpha, P \Rightarrow Q]\}}{G\{\Box\alpha, M \Rightarrow N, [P \Rightarrow Q]\}}\ \Box A_{\mathsf{gl}}$$

$$\frac{G\{M \Rightarrow N, \Box\alpha, [\Box\alpha \Rightarrow \alpha]\}}{G\{M \Rightarrow N, \Box\alpha\}}\ \Box K_{\mathsf{gl}} \qquad \frac{G\{\Box\alpha, M \Rightarrow N, [\Box\alpha, P \Rightarrow Q]\}}{G\{\Box\alpha, M \Rightarrow N, [P \Rightarrow Q]\}}\ 4$$

**Fig. 11.10** The modal rules of the nested sequent system for provability logic GL

e.g., Leivant, 1981; Sambin & Valentin, 1982), cut-free completeness also transfers to the standard sequent system for that logic.

Conversely, using the conversion from standard sequent calculi to nested sequent calculi outlined above, we can decompose the standard sequent rule $GLR$ above into the following nested sequent rules:

$$\frac{G\{\Box\phi, M \Rightarrow N, [\phi, P \Rightarrow Q]\}}{G[\Box\phi, M \Rightarrow N, [P \Rightarrow Q]]}\ \Box L \qquad \frac{G\{\Box\phi, M \Rightarrow N, [\Box\phi, P \Rightarrow Q]\}}{G\{\Box\phi, M \Rightarrow N, [P \Rightarrow Q]\}}\ 4$$

$$\frac{G\{M \Rightarrow N, \Box\phi, [\Box\phi \Rightarrow \phi]\}}{G\{M \Rightarrow N, \Box\phi\}}\ \Box K_{\mathsf{gl}}$$

Note that these are just the nested sequent rules of Fig. 11.10 where in the rule $\Box A_{\mathsf{gl}}$ only the rightmost premiss is kept. Using the simulation of sequent derivations in this calculus as in Proposition 11.5.2, in particular the simulation of the sequent rule for GL by an application of $\Box K_{\mathsf{gl}}$ followed by multiple applications of $\Box L$ and 4, cut-free completeness of the standard sequent system then automatically yields cut-free completeness of this nested sequent system.

### *11.5.2 Hypersequents*

As mentioned above, the problems of the standard sequent framework in capturing prominent modal logics, in particular modal logic S5, early on led to the development of the *hypersequent* framework, introduced independently in Mints (1971), Pottinger (1983) and Avron (1996), and developed extensively by the latter. As mentioned above, in this framework the standard sequent structure is extended to collections of sequents, i.e., lists, sets or multisets, each intuitively corresponding to a separate world in a Kripke-model.

**Definition 11.5.4** A *hypersequent* is a finite multiset of standard sequents, written

$$M_1 \Rightarrow N_1 \mid \ldots \mid M_n \Rightarrow N_n\ .$$

The standard *formula interpretation* of such a hypersequent is given by the formula

$$(M_1 \Rightarrow N_1 \mid \ldots \mid M_n \Rightarrow N_n)^{\Delta} := \Box\left(\bigwedge M_1 \to \bigvee N_1\right) \vee \cdots \vee \Box\left(\bigwedge M_n \to \bigvee N_n\right) .$$

While other formula interpretations have also been considered (see, e.g., Kurokawa, 2014; Lellmann, 2016), for modal logics the one given above is by far the most common. Moreover, in virtually all hypersequent calculi the hypersequent bar | is interpreted as some kind of disjunction. In line with this, essentially all such calculi include some form of the following *external structural rules* of *external weakening*, *external contraction* and *external exchange*:

$$\frac{G}{G \mid M \Rightarrow N} \qquad \frac{G \mid M \Rightarrow N \mid M \Rightarrow N}{G \mid M \Rightarrow N} \qquad \frac{G \mid P \Rightarrow Q \mid M \Rightarrow N}{G \mid M \Rightarrow N \mid P \Rightarrow Q}$$

Note that strictly speaking in our formulation of hypersequents as multisets the external exchange rule is already absorbed into the definition, and hence superfluous. As an example of a hypersequent calculus we consider a calculus for modal logic S5. The most commonly used hypersequent calculus for this logic is the one found in Avron (1996). However, the alternative calculus introduced in Mints (1971) and Poggiolesi (2008a) is better suited to illustrating the relation to nested sequents. The modal rules of the latter calculus are given by[16]:

$$\frac{G \mid M \Rightarrow N, \Box\phi \mid \Rightarrow \phi}{G \mid M \Rightarrow N, \Box\phi} \qquad \frac{G \mid \Box\phi, M \Rightarrow N \mid \phi, P \Rightarrow Q}{G \mid \Box\phi, M \Rightarrow N \mid P \Rightarrow Q} \qquad \frac{G \mid \Box\phi, \phi, M \Rightarrow N}{G \mid \Box\phi, M \Rightarrow N}$$

Note that in this case the modal rules are essentially the modal rules of the nested sequent calculus for modal logic KT with the nesting operator replaced with the hypersequent bar. As explained by Poggiolesi (2010c), this leads to an interpretation of a hypersequent as a nested sequent, where the internal accessibility structure is collapsed into a (multi-)set, i.e., where the accessibility relation is universal.

A slightly different connection between hypersequents and nested sequents is given by the standard formula interpretation $\Delta$ above. Under this interpretation, a hypersequent $M_1 \Rightarrow N_1 \mid \ldots \mid M_n \Rightarrow N_n$ has the same interpretation as the nested sequent $\Rightarrow [M_1 \Rightarrow N_1], \ldots, [M_n \Rightarrow N_n]$. Hence, hypersequents can be seen as a "layer" of a Kripke model, i.e., a number of worlds "side by side" all accessible from a common root. This works reasonably well, e.g., for hypersequent calculi for confluent or linear extensions of modal logic S4. As an example, consider the calculus for modal logic S4.2 (obtained by adding the axiom $\Diamond\Box\phi \to \Box\Diamond\phi$ to an axiomatisation of modal logic S4) from Kurokawa (2014) given by the following characteristic modal rules:

$$\frac{G \mid \phi, M \Rightarrow N}{G \mid \Box\phi, M \Rightarrow N} \qquad \frac{G \mid \Box M \Rightarrow \phi}{G \mid \Box M \Rightarrow \Box\phi} \qquad \frac{G \mid \Box M, \Box N \Rightarrow}{G \mid \Box M \Rightarrow \mid \Box N \Rightarrow}$$

[16] A slightly different version of this hypesequent calculus can be found in Restall (2007).

Informally, the first rule is the standard (local) rule for reflexivity, and the second rule is the standard rule for K4, corresponding to moving from the world represented by the component $\Box M \Rightarrow \Box\phi$ to a successor world, represented by the component $\Box M \Rightarrow \phi$. Taking the interpretation of a hypersequent as a layer of a nested sequent, applying the first rule stays local in a component of the nested sequent, while applying the second advances this component of the layer or slice of the nested sequent one step forward. The third rule, then, informally takes two worlds corresponding to the components $\Box M \Rightarrow$ and $\Box N \Rightarrow$ and replaces them with their common successor world, which corresponds to the component $\Box M, \Box N \Rightarrow$ (recall that the logic is an extension of K4, hence also in this rule the boxes on the left hand side of the sequent arrow carry over to the successor, i.e., premiss of the rule). At this point, however, to present the relations between the components accurately in the nested sequent setting, we would need to introduce a common successor to two nodes in the nested sequent, resulting in an accessibility structure which is no longer a tree (see also Sect. 11.5.6). In general, it seems that the hypersequent framework is perhaps better suited for capturing such "global" properties dealing with two arbitrary worlds in a model and their successors, while the nested sequent framework is based on a more "local" approach and is more suitable for capturing properties which can be expressed by referring to a single world in a Kripke model and its immediate successors.

An intermediate framework between hypersequents and nested sequents is considered under the name of *grafted hypersequents* in Kuznets and Lellmann (2016), where a root node is "grafted" onto a hypersequent, the rules governing the hypersequent layer are those of a hypersequent system, and those governing the connection between the root and the hypersequent layer are those of a nested sequent system.

### *11.5.3 Linear Nested Sequents*

An intermediate framework between standard sequents and nested sequents is given by restricting the tree structure of nested sequents to single branches. Originally introduced in a different notation under the name of *2-sequents* (Masini, 1992), it was also investigated under the names of *non-commutative hypersequents* (Indrzejczak, 2016), *(relational) hypersequents* (Burns & Zach, 2020; Parisi, 2020) or *linear nested sequents*' (Lellmann, 2015; Goré & Lellmann, 2019). Due to its more immediate connection to the nested sequent framework, here we chose to use the latter name. Given that a linear nested sequent is just a branch of a nested sequent, we can equivalently reformulate the linear nested sequent structure as a *list* of sequents:

**Definition 11.5.5** A *linear nested sequent* is a branch of a nested sequent, i.e., a finite list of sequents, written as follows:

$$M_1 \Rightarrow N_1 /\!\!/ \ldots /\!\!/ M_n \Rightarrow N_n$$

The *formula interpretation* of such a linear nested sequent is that of the corresponding nested sequent, i.e., $\bigwedge M_1 \to \bigvee N_1 \vee \Box(\cdots \vee \Box(\bigwedge M_n \to \bigvee N_n)\ldots)$.

Note that the structures of a hypersequent and a linear nested sequent are almost the same, but the formula interpretation is different. In semantic terms, while a hypersequent corresponds to a set or multiset of worlds in a Kripke-model, a linear nested sequent corresponds to a branch of worlds. Since linear nested sequents are restrictions of nested sequents, every linear nested sequent calculus can also be seen as a nested sequent calculus. In fact, e.g., the modal rules of the original calculi from Masini (1992) are exactly the linear versions of the standard nested sequent modal rules for the corresponding logics. Hence completeness immediately transfers from linear nested sequents to nested sequents. If moreover, some care is taken in formulating the nested sequent calculus corresponding to a linear nested sequent calculus, so that in the application of the rules (seen bottom-up) never more than one successor node is created in the nested sequent, and the zero-premiss rules including the initial sequents are restricted to linear nested sequents, then also soundness transfers immediately, and hence the calculi are equivalent. Thus we have:

**Proposition 11.5.6** *Let $\mathcal{L}$ be a logic. If there is a linear nested sequent calculus which is sound and complete for $\mathcal{L}$, then there is also a nested sequent calculus which is sound and complete for $\mathcal{L}$.*

Thus every logic which can be captured by a linear nested sequent calculus can also be captured by a nested sequent calculus, including, e.g., the logics in the above mentioned literature as well as, e.g., the simply dependent multimodal logics considered in Lellmann and Pimentel (2019). Hence the nested sequent framework is at least as strong as the linear nested sequent framework.

For the reverse direction, a set of sufficient criteria for obtaining a linear nested sequent calculus from a nested sequent calculus is given by the above mentioned criteria for converting *basic nested sequent calculi* to standard sequent calculi from Pimentel et al. (2019), see Sect. 11.5.1:

**Proposition 11.5.7** *Let $\mathbf{NS}_{\mathcal{L}}$ be a basic nested sequent calculus which is sound and complete for logic $\mathcal{L}$. Then there is a linear nested sequent calculus which is sound and complete for $\mathcal{L}$.*

***Proof*** It was shown in Pimentel et al. (2019, Theorem 23), that every basic nested sequent calculus can be restricted to its *end-active variant*, i.e., the variant where every rule is restricted so that one of the active components in the premisses is a leaf of the nested sequent. Since, moreover, by definition every premiss of a modal basic nested sequent rule only mentions one nesting operator, we can restrict the end-active variant of $\mathbf{NS}_{\mathcal{L}}$ to linear nested sequents, simply by restricting all rules to linear nested sequents. Call the resulting linear nested sequent calculus $\mathsf{LNS}_{\mathcal{L}}$. Since every derivation in $\mathsf{LNS}_{\mathcal{L}}$ is also a derivation in $\mathbf{NS}_{\mathcal{L}}$, soundness follows immediately. To show completeness of $\mathsf{LNS}_{\mathcal{L}}$ we simply follow the completeness proof for the corresponding standard sequent system in Sect. 11.4 of *op. cit.* by rearranging nested

sequent derivations in the end-active variant of $\mathbf{NS}_{\mathcal{L}}$ into blocks consisting of a successor creating (*creation*) rule followed by upgrade rules to populate the newly created successor, and restricting the nested sequents to linear nested sequents using a suitable form of the disjunction property (Lem. 28 in *op. cit.*). □

While the form of a basic nested sequent system is sufficient for restricting to a linear nested sequent system, it is not necessary: In particular they are not satisfied by the linear nested sequent calculi for linear tense logics of Indrzejczak (2016) and those for modal logic $\mathsf{KB}$ and modal tense logic $\mathsf{K}_t$ from Goré and Lellmann (2019), and hence by the corresponding nested sequent calculi. It is still an open problem to formulate more general criteria for restricting nested sequent calculi to linear nested sequent calculi, which capture systems for logics like the ones above which do not have a cut-free sequent calculus already.

### *11.5.4 Prefixed Tableaux*

In classical logic as well as in modal logic, it is a well-known fact that standard sequent calculi correspond to *semantic tableaux calculi* (see, e.g., Smullyan, 1968; Fitting, 1983). This correspondence can be easily explained. For the technical details, see *op. cit.* The idea is to try to construct a countermodel for (the formula interpretation of) a given sequent by applying the semantic tableaux rules (see below for details) backward until we reach an initial sequent (a *closed branch* in tableaux terminology) or every possible rule has already been applied, possibly taking into account some loop checking conditions (an *open branch*). Since the emphasis is on the countermodel construction, the presentation of sequents is also more semantic: A sequent $\phi_1, \ldots, \phi_n \Rightarrow \epsilon_1, \ldots, \epsilon_m$ is replaced with the *signed formulae* $\mathsf{T} : \phi_1, \ldots, \mathsf{T} : \phi_n, \mathsf{F} : \epsilon_1, \ldots, \mathsf{F} : \epsilon_m$, corresponding to the fact that in a countermodel for the interpretation $\phi_1 \wedge \cdots \wedge \phi_n \to \epsilon_1 \vee \cdots \vee \epsilon_m$ of this sequent, the formulae $\phi_1, \ldots, \phi_n$ on the left hand side of the sequent must be true, whereas the formulae $\epsilon_1, \ldots, \epsilon_m$ on the right hand side must be false. *Branch extension rules*, which specify how to deconstruct the principal formula of the sequent rule and add its components to a branch, correspond to sequent rules. E.g., the (classical, Kleene'd) conjunction left and right rules in the sequent framework and their corresponding tableaux rules are:

$$\frac{M, \phi \wedge \epsilon, \phi, \epsilon \Rightarrow N}{M, \phi \wedge \epsilon \Rightarrow N} \wedge_L \qquad \frac{\mathsf{T} : \phi \wedge \epsilon}{\begin{array}{c}\mathsf{T} : \phi \\ \mathsf{T} : \epsilon\end{array}}$$

$$\frac{M \Rightarrow N, \phi \wedge \epsilon, \phi \quad M \Rightarrow N, \phi \wedge \epsilon, \epsilon}{M \Rightarrow N, \phi \wedge \epsilon} \wedge_R \qquad \frac{\mathsf{F} : \phi \wedge \epsilon}{\mathsf{F} : \phi \mid \mathsf{F} : \epsilon}$$

The way to read tableaux rules is that the rule corresponding to the conjunction left rule $\wedge_L$ can be applied to a branch containing the signed formula $\mathsf{T} : \phi \vee \epsilon$ by

adding both signed formulae $\mathsf{T} : \phi$ and $\mathsf{T} : \epsilon$ to the bottom of this branch. The rule corresponding to the conjunction right rule $\wedge_R$ can be applied to a branch containing the signed formula $\mathsf{F} : \phi \wedge \epsilon$ by splitting it into two branches at the bottom, one with the additional signed formula $\mathsf{F} : \phi$, the other with $\mathsf{F} : \epsilon$. Note that we considered the Kleene's version of the sequent rules, i.e., the version where the principal formula is copied into the premisses, to correspond to the fact that signed formulae are added to a branch of a tableau without removing the original signed formula. Branches in the process of constructing a tableau thus correspond directly to sequents at the corresponding stage of backward proof search in the sequent system. In particular, a branch is *closed* if for some formula $\delta$ it contains both $\mathsf{T} : \delta$ and $\mathsf{F} : \delta$, and hence closed branches correspond to initial sequents with the formula $\delta$ occurring both on the left and right hand side of the sequent. A tableau starting from a single node containing only a formula $\mathsf{F} : \delta$ in which every branch is closed thus corresponds to a sequent derivation of the sequent $\Rightarrow \delta$. Conversely, given the direct connection of formula signs to truth values, a countermodel can be read off immediately from an open branch to which every possible rule (modulo loop checking) has been applied by assigning every variable in the branch the truth value corresponding to its sign.

Where semantic tableaux calculi can be seen as the upside-down version of standard sequent calculi, the analogous upside-down version of nested sequent calculi is given by *prefixed tableaux* calculi, a framework going back already to Fitting (1972). The correspondence was first shown in Fitting (2012) based on the single-sided formulation of nested sequent calculi. To align better with the two-sided formulation used here we adapt the two-sided notation from Fitting (2014), used there for intuitionistic logic. The main idea is to capture the tree-structure of nested sequents and Kripke-models in an additional prefix to every formula, essentially a name for the component of the nested sequent which contains this formula and the corresponding world in a Kripke-model. Formally, a *prefix* is a non-empty and finite sequence of positive numbers starting with 1, written using periods as separators. Examples are given by the prefixes 1.1.2.1 and 1.2.3.2. A set of prefixes which is closed under taking initial segments thus yields a tree-structure: the successors of a prefix $\sigma$ are the prefixes $\sigma.n$ for positive integers $n$. A *prefixed signed formula* then has the form $\sigma : \mathsf{T} : \phi$ or $\sigma : \mathsf{F} : \phi$ for some prefix $\sigma$ and formula $\phi$. Using the ideas of the classical case but extended with prefixes as names for components of a nested sequent, we can then convert back and forth between nested sequent calculi and prefixed tableaux calculi. E.g., the modal rules of the Kleene'd calculus $\mathbf{NS}^+_{\mathsf{k}}$ for modal logic $\mathsf{K}$ from Sect. 11.3 together with the corresponding prefixed tableaux rules are as follows:

$$\frac{G\{\Box\phi, M \Rightarrow N, [\phi, P \Rightarrow Q]\}}{G\{\Box\phi, M \Rightarrow N, [P \Rightarrow Q]\}}\ \Box L^+ \qquad \frac{\sigma : \mathsf{T} : \Box\phi}{\sigma.n : \mathsf{T} : \phi}\ \sigma.n \text{ not new}$$

$$\frac{G\{M \Rightarrow N, \Box\phi, [\Rightarrow \phi]\}}{G\{M \Rightarrow N, \Box\phi\}}\ \Box R^+ \qquad \frac{\sigma : \mathsf{F} : \Box\phi}{\sigma.n : \mathsf{F} : \phi}\ \sigma.n \text{ new}$$

Here the additional restrictions on the prefix state that in the rule corresponding to $\Box L^+$ the prefix $\sigma.n$ is *not new* on the branch, i.e., there is a formula $\sigma.n : \mathsf{F} : \delta$ or $\sigma.n : \mathsf{T} : \delta$ already on the branch, corresponding to the fact that the component $[P \Rightarrow Q]$ exists in the nested sequent. In contrast, for the prefixed tableau rule corresponding to the nested sequent rule $\Box R^+$ the prefix $\sigma.n$ must be *new* in the branch, corresponding to the fact that the component $[\Rightarrow \phi]$ is freshly created by the nested sequent rule.

In the same spirit, all the logical variants of the nested sequent calculi from Sect. 11.2 can be seen essentially as notational variants of prefixed tableaux calculi and vice versa. The details of the proof for equivalence of the (single-sided) nested sequent calculi and the corresponding prefixed tableaux calculi are given in Fitting (2012). As also noted in *op. cit.*, under this connection the extensive body of work on prefixed tableaux thus gives rise to a wealth of nested sequent calculi. See, e.g., Massacci (2000), Castilho et al. (1997) or Goré (1999) for an overview.

### *11.5.5 Labelled Sequents*

A different semantically motivated extension of the sequent framework is given by the framework of *labelled sequents*, investigated prominently, e.g., in Simpson (1994), Viganò (2000), Negri (2005) and Miller and Volpe (2015). Here, similarly to the framework of prefixed tableaux, formulae are prefixed with a *label* corresponding to the world in the Kripke-model in which the formula is supposed to be true/false. The difference is however, that the relational structure of the model is captured explicitly in the language. This is achieved by the use of *relational atoms*, i.e., formulae $x\,R\,y$ for $x$ and $y$ labels, which are interpreted as explicitly stating that the world denoted by $y$ is a successor of the world denoted by $x$. Using this explicit representation, a wide ranging class of semantical properties can be captured and the corresponding logics automatically are endowed with a cut-free sequent system (see *op. cit.*). Compared to the tree structure used in nested sequents, the structure given by an arbitrary relation on the worlds is of course much more general. The subclass of labelled sequents corresponding to nested sequents was characterised in Goré and Ramanayake (2012) by restricting the relational atoms to spell out the structure of a tree:

**Definition 11.5.8** A *labelled tree sequent* is a sequent $\mathcal{R}, M \Rightarrow N$ where

- $\mathcal{R}$ is a *treelike relation set*, i.e., a set of relational atoms of the form $x\,R\,y$ which is empty or such that the induced relation on the set of variables occurring in $\mathcal{R}$ is a tree
- $M$ and $N$ consist of *labelled formulae*, i.e., formulae of the form $x : \phi$ for a world variable or *label* $x$
- if $\mathcal{R} = \emptyset$, then all formulae in $M \Rightarrow N$ have the same label
- if $\mathcal{R} \neq \emptyset$, then every label occurring in a formula in $M \Rightarrow N$ also occurs in $\mathcal{R}$.

Similarly to the prefixes in the prefixed tableaux framework, the labels of formulae can then be taken as names for components of a nested sequent, with the successor

relation in the nested sequent corresponding to the relation induced by the relation set. The restriction to treelike relation sets ensures that the resulting structure is a tree and hence a nested sequent. Thus as shown in *op. cit.*, nested sequent calculi can be translated into labelled tree sequent calculi and vice versa.

As an example, the modal nested sequent rules of the nested sequent calculus $\mathbf{NS}_{\mathsf{K}}$ (see Fig. 11.4) below left correspond to the labelled tree sequent rules below right:

$$\frac{G\{\Box\phi, M \Rightarrow N, [\phi, P \Rightarrow Q]\}}{G\{\Box\phi, M \Rightarrow N, [P \Rightarrow Q]\}}\,\Box L \qquad \frac{\mathcal{R}, xRy, X, x : \Box\phi, y : \phi \Rightarrow Y}{\mathcal{R}, xRy, X, x : \Box\phi \Rightarrow Y}$$

$$\frac{G\{M \Rightarrow N, [\,\Rightarrow \phi]\}}{G\{M \Rightarrow N, \Box\phi\}}\,\Box R \qquad \frac{\mathcal{R}, xRy, X \Rightarrow Y, y : \phi}{\mathcal{R}, X \Rightarrow Y, x : \Box\phi}$$

Here the label $y$ in the rule corresponding to $\Box R$ is a name for the new component $[\,\Rightarrow \phi]$, and hence must not occur in the conclusion of that rule. The label $x$ in that rule is the name for the component $[M \Rightarrow N, \Box\phi]$, and the relational atom $xRy$ specifies that the component $[\,\Rightarrow \phi]$ is one of its successors. Note that these are exactly the rules of the labelled sequent calculus **G3K** for modal logic $\mathsf{K}$ from Negri (2005). Thus, as in the case of prefixed tableaux, the two calculi can be seen as notational variants (see also Poggiolesi 2010a, b, Sect. 5.4 ). Similarly, the logical variants of the calculi from Sect. 11.2 can be seen as labelled tree sequent calculi. As it was shown in Goré and Ramanayake (2012) using the example of provability logic, this correspondence between nested sequent calculi and labelled tree sequent calculi automatically transfers results such as cut-free completeness. It should be pointed out, that the labelled sequent rules corresponding directly to the standard nested sequent rules for extensions of $\mathsf{K}$ do not match the standard labelled sequent rules for these logics. An emulation of these nested sequent calculi in a focused version of the standard labelled sequent calculi for these logics is given in Marin et al. (2016).

As with prefixed tableaux, certain structural rules like the rule $\tilde{b}$ are easier formulated in the nested sequent framework, since there it is not necessary to simultaneously rename all labels in a sequent. For labelled tree sequents this however works slightly better than for prefixed tableaux, since the tree structure is local to a sequent and not distributed along a branch. Hence this structure can be changed by modifying only a single sequent instead of potentially all formulae along the branch of a tableau down to the root. The latter operation in particular might result in conflicts with the other branches of the tableau.

Conversely, in the labelled setting it is very natural to capture properties of the models using rules which only work on the relation sets, and to keep the rules working on formulae of the form $x : \Box\phi$ constant. E.g., the rules for transitivity and reflexivity are naturally formulated as:

$$\frac{\mathcal{R}, xRy, yRz, xRz, X \Rightarrow Y}{\mathcal{R}, xRy, yRz, X \Rightarrow Y}\ \mathsf{trans} \qquad \frac{\mathcal{R}, xRx, X \Rightarrow Y}{\mathcal{R}, X \Rightarrow Y}\ \mathsf{refl}$$

Since such rules very easily destroy the tree structure of the relation set, converting labelled sequent calculi with such rules into labelled tree sequent calculi and hence nested sequent calculi is often impossible. Where it is possible, the conversion is often rather involved and non-trivial, see, e.g., Simpson (1994) and Lyon (2020a, b). For a more thorough overview and an exploration of a rather general method for such a conversion, see also the recent (Lyon, 2021b).

### *11.5.6 Indexed Nested Sequents*

As mentioned above, nested sequents are trees of sequents, and hence when seen semantically essentially correspond to Kripke models based on trees. Due to this close connection to trees, the nested sequent formalism tends to struggle with modal logics characterised by frame properties which require a different structure, such as *confluence*, stating that for all worlds $x, y, z$ in a Kripke frame with $xRy$ and $xRz$ there is a common successor $w$ with $xRw$ and $yRw$. To alleviate this problem, nested sequents were generalized in Fitting (2015) to *indexed nested sequents*.[17] Intuitively, every component in a nested sequent is extended with an *index*, loosely speaking a name for the corresponding world in a Kripke model. In particular, different components with the same index correspond to the same world of a model. This way it is possible to mirror properties like confluence in the syntactic framework by adding successors to two components in the tree structure of a nested sequent, and identifying the successors by giving them the same index. To keep in line with the two-sided notation in this article, we follow the notation from Ramanayake (2018).

**Definition 11.5.9** An *indexed sequent* is obtained by adding an *index* $n \in \mathbb{N}$ to a sequent $M \Rightarrow N$, written $M \overset{n}{\Rightarrow} N$. An *indexed nested sequent* is of the form

$$M \overset{n}{\Rightarrow} N, [G_1], \ldots, [G_m]$$

where $n$ is an index, $m \geq 0$ and $G_1, \ldots, G_m$ are indexed nested sequents.

Other notions like that of a *context* then are defined analogously to the nested sequent framework. The modal rules of an indexed nested sequent system for normal modal logic K then are essentially the nested sequent rules with additional indices:

$$\frac{G\{\Box\phi, M \overset{a}{\Rightarrow} N, [\phi, P \overset{b}{\Rightarrow} Q]\}}{G\{\Box\phi, M \overset{a}{\Rightarrow} N, [P \overset{b}{\Rightarrow} Q]\}}\ \Box L \qquad \frac{G\{M \overset{a}{\Rightarrow} N, \Box\phi, [\overset{b}{\Rightarrow} \phi]\}}{G\{M \overset{a}{\Rightarrow} N, \Box\phi\}}\ \Box R$$

[17] To prevent confusion about the terminology: Structures called *indexed nested sequents* were also independently introduced in Poggiolesi (2013). However in Poggiolesi's terminology, they denote nested sequents with indices for agents, and hence do not break the tree structure of nested sequents. They are used to capture proof-theoretically multi-agent epistemic logic. In Bull (1992), nested sequents have also been indexed with programs. See Sect. 11.4.3.

where the rule $\Box R$ has the side condition that the index $b$ does not occur in the conclusion. However, since two components with the same index correspond to the same world in the Kripke model, the calculus also includes rules to communicate between such components:

$$\frac{G\{\phi, M \overset{a}{\Rightarrow} N\}\{\phi, P \overset{a}{\Rightarrow} Q\}}{G\{\phi, M \overset{a}{\Rightarrow} N\}\{P \overset{a}{\Rightarrow} Q\}}\ \mathsf{fc-l} \qquad \frac{G\{M \overset{a}{\Rightarrow} N, \phi\}\{P \overset{a}{\Rightarrow} Q, \phi\}}{G\{M \overset{a}{\Rightarrow} N, \phi\}\{P \overset{a}{\Rightarrow} Q\}}\ \mathsf{fc-r}$$

$$\frac{G\{M \overset{a}{\Rightarrow} N, [P \overset{b}{\Rightarrow} Q]\}\{R \overset{a}{\Rightarrow} S, [\overset{b}{\Rightarrow}]\}}{G\{M \overset{a}{\Rightarrow} N, [P \overset{b}{\Rightarrow} Q]\}\{R \overset{a}{\Rightarrow} S\}}\ \mathsf{sc}$$

Here the notation $G\{\ \}\{\ \}$ denotes the obvious extension of an indexed nested sequent with a context to an indexed nested sequent with *two* (possibly identical) contexts. We refer to, e.g., Ramanayake (2018) for the full definition. Intuitively, the rules $\mathsf{fc-l}$ and $\mathsf{fc-r}$ state that formulae can move freely between components with the same index. Rule $\mathsf{sc}$ intuitively states that a component with index $a$ has the same successors, irrespective of where in the indexed nested sequent it occurs.

The frame property of confluence mentioned above then is captured by the following additional rule:

$$\frac{G\{M \overset{a}{\Rightarrow} N, [P \overset{b}{\Rightarrow} Q, [\overset{d}{\Rightarrow}]], [R \overset{c}{\Rightarrow} S, [\overset{d}{\Rightarrow}]]\}}{G\{M \overset{a}{\Rightarrow} N, [P \overset{b}{\Rightarrow} Q], [R \overset{c}{\Rightarrow} S]\}}$$

This rule rather directly captures the frame condition that if the worlds corresponding to the indices $b$ and $c$ are successors of the one corresponding to $a$, then they have a common successor world corresponding to the index $d$. More generally, the indexed nested sequent formalism was used in Fitting (2015) to introduce calculi for modal logics axiomatised by *Geach formulae*, also known as *Scott–Lemmon axioms*, i.e., formulae of the form $\Diamond^k\Box^\ell\phi \to \Box^m\Diamond^n\phi$ for $k, \ell, m, n \geq 0$. Syntactic completeness for these calculi via direct cut elimination and their extension to the intuitionistic setting were investigated in Marin and Straßburger (2017).

The precise relation of indexed nested sequents to nested sequents is clarified in Ramanayake (2018) via a correspondence to a restricted form of labelled sequents (see Sect. 11.5.5 above). In particular, in *op. cit.* it was shown that indexed nested sequents correspond to *labelled tree sequents with equality*, an extension of the labelled tree sequents of Definition 11.5.8 to labelled sequents of the form $\mathcal{R}, \mathcal{E}, M \Rightarrow N$ where $\mathcal{R}$ again is a tree-like relation set, $M$ and $N$ contain labelled formulae, and $\mathcal{E}$ now contains *equality terms* of the form $x = y$. Again, labelled tree sequents with equality are subject to certain restrictions ensuring that every variable occurring in $M, N$ or $\mathcal{E}$ occurs also in $\mathcal{R}$ if the latter is non-empty. In case $\mathcal{R}$ is empty, every variable occurring in $M, N$ or $\mathcal{E}$ is the same. Thus, the difference in expressive

power between the frameworks of nested sequents and indexed nested sequents is essentially the difference in expressive power between the frameworks of labelled tree sequents and labelled tree sequents with equality.

### *11.5.7 Display Calculi*

Considering only the syntactic shape of nested sequents, it might seem a bit strange that the nesting operator is allowed to occur only on the right hand side of the sequent arrow. After all, the comma as a structural connective might appear on both sides. Moreover, while in the nested sequent framework we have structural connectives corresponding to implication (the sequent arrow), conjunction (the comma on the left), disjunction (the comma on the right), and the modal box (the nesting operator), there are no structural connectives corresponding to negation $\neg$, falsum $\bot$ or verum $\top$. So it would seem a natural step to consider sequent structures which permit the nesting operator [.] on both sides of the sequent arrow and include structural connectives for every logical connective in the language. In fact, this is essentially what happens in the framework of *display calculi*, introduced in Belnap (1982). While adding structural connectives for the propositional connectives works rather smoothly, though, in order to give a formula interpretation for the nesting operator on the left hand side of such structures we need to extend the language to that of *tense logics* (see also Sect. 11.4.3), by introducing the modal connective $\blacklozenge$ for the residual of the modal box (see Wansing, 2002). Semantically, this is interpreted as a backward looking diamond: a formula $\blacklozenge\phi$ is true in a world of a Kripke frame if the formula $\phi$ is true in at least one of its predecessors.

**Definition 11.5.10** A *(modal) display sequent* is a tuple $X \Rightarrow Y$, where $X$ and $Y$ are *display structures* given by the following grammar:

$$X ::= \phi \mid \mathbf{I} \mid *X \mid \bullet X \mid (X \circ X)$$

where $\phi$ is a formula of modal tense logic. The interpretation of the left and right hand sides of a display sequent is given by the two functions $\ell$ and $r$ as follows:

$$\begin{array}{rlrl} \ell(\phi) &= \phi & r(\phi) &= \phi \\ \ell(\mathbf{I}) &= \top & r(\mathbf{I}) &= \bot \\ \ell(*X) &= \neg r(X) & r(*X) &= \neg \ell(X) \\ \ell(\bullet X) &= \blacklozenge \ell(X) & r(\bullet X) &= \Box r(X) \\ \ell(X \circ Y) &= \ell(X) \wedge \ell(Y) & r(X \circ Y) &= r(X) \vee r(Y) \end{array}$$

The formula interpretation of a display sequent $X \Rightarrow Y$ is $\ell(X) \rightarrow r(Y)$.

Note that the comma of standard or nested sequents is replaced with a binary structural connective $\circ$, making properties like associativity of conjunction/disjunction explicit. This has the benefit that the logical rules are applied only to an isolated or

*displayed* formula. For example, the modal rules of a display calculus for tense logic $\mathsf{K}_t$, the tense analogue of modal logic $\mathsf{K}$ are:

$$\frac{\phi \Rightarrow X}{\Box\phi \Rightarrow \bullet X}\ \Box_L \qquad \frac{X \Rightarrow \bullet\phi}{X \Rightarrow \Box\phi}\ \Box_R \qquad \frac{\bullet\phi \Rightarrow X}{\blacklozenge\phi \Rightarrow X}\ \blacklozenge_L \qquad \frac{X \Rightarrow \phi}{\bullet X \Rightarrow \blacklozenge\phi}\ \blacklozenge_R$$

However, in order to be able to display a formula so that a logical rule can be applied to it, display calculi require the addition of *display rules*. For example, for the structural connective $\bullet$ these are

$$\frac{X \Rightarrow \bullet Y}{\bullet X \Rightarrow Y} \qquad \frac{\bullet X \Rightarrow Y}{X \Rightarrow \bullet Y} \qquad \frac{\mathbf{I} \Rightarrow Y}{\bullet\mathbf{I} \Rightarrow Y} \qquad \frac{X \Rightarrow \mathbf{I}}{X \Rightarrow \bullet\mathbf{I}}$$

A detailed introduction to display calculi including the display rules for all the structural connectives is beyond the scope of this article—we refer the reader to, e.g., Belnap (1982) and Wansing (1998). From the perspective of nested sequents the two display rules above left are interesting, though. Intuitively, while the operator $\bullet$ on the right hand side of a display sequent corresponds to the nesting operator [.] in a nested sequent, and hence to moving to a successor node in the underlying tree, the operator $\bullet$ on the left hand side of a display sequent corresponds to moving to a predecessor node. Hence, moving between premiss and conclusion in the display rules essentially corresponds to moving between different nodes in a Kripke-model. Viewing a Kripke-model as a nested sequent, then, intuitively, the display rules can be used to shift the root of a nested sequent to a different node, so that logical rules can be applied to formulae there. This intuition is in fact already present in Kashima (1994), one of the works which introduced the nested sequent framework (in the single-sided formulation, the display rules above correspond to the *turn* rules introduced in *op. cit.*). Using this is it possible to simulate nested sequent derivations using the display rules. e.g., the application of the nested sequent $\Box_R$ rule below left in a successor of the root node can be simulated by the display derivation below right, where the applications of the display rules in the first and last step correspond to shifting the root node so that the logical rule $\Box_R$ can be applied.

$$\frac{\phi \Rightarrow [\ \Rightarrow [\ \Rightarrow \epsilon\ ]]}{\phi \Rightarrow [\ \Rightarrow \Box\epsilon\ ]}\ \Box_R \qquad \dfrac{\dfrac{\dfrac{\phi \Rightarrow \bullet\bullet\epsilon}{\bullet\phi \Rightarrow \bullet\epsilon}}{\bullet\phi \Rightarrow \Box\epsilon}\ \Box_R}{\phi \Rightarrow \bullet\Box\epsilon}$$

Making this connection formally precise is somewhat cumbersome and heavily relies on the display rules for the remaining connectives. Hence we refer the reader to Poggiolesi (2010a) for the details.

Given the previous remarks it should be clear that display calculi are at least as expressive as nested sequent calculi. Since there are very general results about how to automatically obtain cut free display calculi for logics given by a large class of axioms (see, e.g., Kracht, 1996), it is then of course interesting to see which kinds of display

calculi can be converted into nested sequent calculi. Some very general results in this spirit, albeit based on slightly different display-style calculi, are shown in Goré et al. (2011), where the authors use a transformation from display-like *shallow inference* calculi to nested sequent (or *deep inference*) calculi to uniformly obtain cut-free modular systems for modal tense logics given by *Scott–Lemmon axioms* (see also Sect. 11.5.6) or by *path axioms*, i.e., axioms of the form $\langle\rangle_1 \ldots, \langle\rangle_n \phi \rightarrow \langle\rangle \phi$, where $n \geq 0$ and each of $\langle\rangle_1, \ldots, \langle\rangle_n, \langle\rangle$ is either $\Diamond$ or $\blacklozenge$. A further extension of these results is presented in the very recent (Ciabattoni et al., 2021).

It is worth noting, however, that the greatly increased expressivity of display calculi compared to nested sequent calculi comes at a price: First, owing to the fact that the presence of the display rules essentially necessitates working with equivalence classes of display sequents, proof search in display calculi is a notoriously difficult problem, rendering such calculi subobtimal for decidability or complexity results. Moreover, the display rules for the nesting operator $\bullet$ as well as the formula interpretation for this operator on the left hand side heavily rely on the presence of the residuum $\blacklozenge$ for the modal operator $\Box$ in the language. While in the modal case often it can be shown that the display calculi are conservative for the fragment of the language not containing the operator $\blacklozenge$, in general this can be complicated, and for some logics such as non-normal modal logics it might also not be clear what the residuum for the modal box is.

## 11.6 Conclusion

Modal logic is the logic that was originally introduced to treat the concepts of necessity and possibility. While its basic principles were the subject of study and debate already since the time of Aristotle, it is to a large extent thanks to the introduction of Kripke semantics, in the early 1960s, that modal logic has become a powerful, fruitful and very much used non-classical logic. However, given the difficulties encountered in the search for cut-free sequent calculi for various important modal logics such as S5, until the end of the last century there was a general skepticism about the possibility of a thriving general proof theory for modal logic. The tree-hypersequent, deep sequent or nested sequent method is amongst the methods that helped change this perspective. The main aim of this paper has been to offer a reasonably complete, faithful and harmonious panorama of nested sequents and associated results. After having introduced nested sequents, we have shown the main theorems that can be obtained with it, e.g., cut-elimination, decidability and interpolation; moreover, we have illustrated its relation with other well-known proof-theoretic formalisms and outlined the many logics it can be applied to. With this we encourage the reader to dive deeper into the provided references and hope to contribute to the study of nested sequent calculi as a vibrant and growing area of research.

## References


Alberucci, L., & Jager, G. (2005). About cut elimination for logics of common knowledge. *Annals of Pure and Applied Logic, 133*, 73–99.

Alenda, R., Olivetti, N., & Pozzato, G. L. (2013). Nested sequent calculi for normal conditional logics. *Journal of Logic and Computation, 26*(1), 7–50. https://doi.org/10.1093/logcom/ext034

Arisaka, R., Das, A., & Strassburger, L. (2015). On nested sequents for constructive modal logics. *Logical Methods in Computer Science*, *11*(3). https://doi.org/10.2168/LMCS-11(3:7)2015

Artemov, S. (2008). The logic of justification. *The Review of Symbolic Logic, 4*, 477–513.

Avron, A. (1996). The method of hypersequents in the proof theory of propositional non-classical logic. In W. Hodges, M. Hyland, C. Steinhorn, & J. Strauss (Eds.), *Logic: From Foundations to Applications* (pp. 1–32). Oxford: Oxford University Press.

Belnap, N. D. (1982). Display logic. *Journal of Philosophical Logic, 11*, 375–417.

Bílková, M. (2011). A note on uniform interpolation proofs in modal deep inference calculi. In: N. Bezhanishvili, S. Löbner, K. Schwabe, L. Spada (Eds.), *Logic, language, and computation. TbiLLC 2009. Lecture notes in computer science* (Vol. 6618, pp. 30–45). Berlin: Springer. https://doi.org/10.1007/978-3-642-22303-7_3

Blackburn, P., de Rijke, M., & Venema, Y. (2001). *Modal logic*. Cambridge: Cambridge University Press.

Blamey, S., & Humberstone, L. (1991). A perspective on modal sequent logic. *Publications of the Research Institute for Mathematical Sciences, Kyoto University, 27*, 763–782.

Brown, M. A. (1988). On the logic of ability. *Journal of Philosophical Logic, 17*(1), 1–26.

Brünnler, K. (2006). Deep sequent systems for modal logic. In G. Governatori, I. Hodkinson, & Y. Venema (Eds.), *Advances in modal logic* (Vol. 6, pp. 107–119). London: College Publications.

Brünnler, K. (2009). Deep sequent systems for modal logic. *Archive for Mathematical Logic, 48*, 551–577.

Brünnler, K. (2010). How to universally close the existential rule. In C. Fermüller & A. Voronkov (Eds.), *Logic for programming, artificial intelligence, and reasoning. LPAR 2010. Lecture notes in computer science* (Vol. 6397, pp. 172–186). Berlin: Springer. https://doi.org/10.1007/978-3-642-16242-8_13

Brünnler, K., & Stüder, T. (2009). Syntactic cut-elimination for common knowledge. *Annals of Pure and Applied Logic, 160*, 82–95.

Bull, R. A. (1992). Cut elimination for propositional dynamic logic without *. *Mathematical Logic Quarterly, 38*, 85–100.

Burns, S., & Zach, R. (2020). Cut-free completeness for modular hypersequent calculi for modal logics K, T, and D. *Review of Symbolic Logic,* 1–20. https://doi.org/10.1017/S1755020320000180

Castilho, M. A., del Cerro, L. F., Gasquet, O., & Herzig, A. (1997). Modal tableaux with propagation rules and structural rules. *Fundamenta Informaticae, 32*, 281–297.

Cerrato, C. (1993). Cut-free modal sequents for normal modal logics. *Notre Dame Journal of Formal Logic, 34*, 564–582.

Chellas, B. F. (1980). *Modal logic*. Cambridge University Press.

Ciabattoni, A., Lyon, T. S., Ramanayake, R., & Tiu, A. (2021). Display to labeled proofs and back again for tense logics. *ACM Transactions on Computational Logic*, *22*(3), 1–31. https://doi.org/10.1145/3460492

Curry, H. B. (1952). The elimination theorem when modality is present. *Journal of Symbolic Logic, 17*, 249–265.

Day, A., & Jaroslav, J. (1984). The amalgamation property for varieties of lattices. *Transactions of the American Mathematical Society*, *286*, 251–266.

Dragalin, A. G. (1988). *Mathematical intuitionism. Introduction to proof theory*. Providence: American Mathematical Society.

Fagin, R., Halpern, Y., & Vardi, M. (2004). *Reasoning about knowledge*. MIT Press.

Fitting, M. (1972). Tableau methods of proof for modal logics. *Notre Dame Journal of Formal Logic, 13*, 237–247.

Fitting, M. (1983). *Proof methods for modal and intuitionistic logics. Synthese library (SYLI)* (Vol. 169). Berlin: Springer. https://doi.org/10.1007/978-94-017-2794-5

Fitting, M. (2012). Prefixed tableaus and nested sequents. *Annals of Pure and Applied Logic, 163*(3), 291–313.

Fitting, M. (2014). Nested sequents for intuitionistic logics. *Notre Dame Journal of Formal Logic, 55*(1), 41–61. https://doi.org/10.1215/00294527-2377869

Fitting, M. (2015). Cut-free proof systems for Geach logics. *IfCoLog Journal of Logics and Their Applications, 2*(2), 17–64.

Fitting, M., & Kuznets, R. (2015). Modal interpolation via nested sequents. *Annals of Pure and Applied Logic, 166*, 274–305.

Galmiche, D., & Salhi, Y. (2010). Label-free natural deduction systems for intuitionistic and classical modal logics. *Journal of Applied Non-Classical Logics, 20*(4), 373–421.

Galmiche, D., & Salhi, Y. (2018). Tree-sequent calculi and decision procedures for intuitionistic modal logics. *Journal of Logic and Computation, 28*(5), 967–989. https://doi.org/10.1093/logcom/exv039

Gentzen, G. (1935). Untersuchungen über das logische Schließen. I. *Mathematische Zeitschrift, 39*(1), 176–210.

Girlando, M., Lellmann, B., & Olivetti, N. (2019). Nested sequents for the logic of conditional belief. In: F. Calimeri, N. Leone, M. Manna (Eds.), *Logics in artificial intelligence. JELIA 2019. Lecture notes in computer science* (Vol. 11468, pp. 709–725). Berlin: Springer. https://doi.org/10.1007/978-3-030-19570-0_46

Girlando, M., & Straßburger, L. (2020). MOIN: A nested sequent theorem prover for intuitionistic modal logics (system description). In N. Peltier & V. Sofronie-Stokkermans (Eds.), *Automated reasoning. IJCAR 2020. Lecture notes in computer science* (Vol. 12167, pp. 398–407). Cham: Springer. https://doi.org/10.1007/978-3-030-51054-1_25

Goble, L. F. (1974). Gentzen systems for modal logic. *Notre Dame Journal of Formal Logic, 15*(3), 455–461.

Goré, R. (1999). Tableau methods for modal and temporal logics. In M. D'Agostino, D. M. Gabbay, R. Hähnle, & J. Posegga (Eds.), *Handbook of tableau methods* (pp. 297–396). Dordrecht: Kluwer.

Goré, R., & Lellmann, B. (2019). Syntactic cut-elimination and backward proof-search for tense logic via linear nested sequents. In S. Cerrito, A. Popescu (Eds.), *Automated reasoning with analytic tableaux and related methods. TABLEAUX 2019. Lecture notes in computer science* (Vol. 11714, pp. 182–202). Berlin: Springer. https://doi.org/10.1007/978-3-030-29026-9_11

Goré, R., Postniece, L., & Tiu, A. (2008). Cut-elimination and proof-search for bi-intuitionistic logic using nested sequents. In *AiML* 2008.

Goré, R., Postniece, L., & Tiu, A. (2011). On the correspondence between display postulates and deep inference in nested sequent calculi for tense logics. *Log. Methods Comput. Sci.*, *7*(2:8), 1–38.

Goré, R., & Ramanayake, R. (2012). Labelled tree sequents, tree hypersequents and nested (deep) sequents. In T. Bolander, T. Braüner, S. Ghilardi, & L. S. Moss (Eds.), *AiML 9* (pp. 279–299). College Publications.

Guerrini, S., Martini, S., & Masini, A. (1998). An analysis of (linear) exponentials based on extended sequents. *Logic Journal of the IGPL, 6*(5), 735–753. https://doi.org/10.1093/jigpal/6.5.735

Hill, B., & Poggiolesi, F. (2010). A contraction-free and cut-free sequent calculus for propositional dynamic logic. *Studia Logica, 94*, 68–94.

Hill, B., & Poggiolesi, F. (2015). Common knowledge: finite calculus with syntactic cut-elimination procedure. *Logique et Analyse, 58*, 136–159.

Hill, B., & Poggiolesi, F. (2019). An analytic calculus for the intuitionistic logic of proofs. *Notre Dame Journal of Formal Logic, 60*, 353–393.

Hodges, W. (1997). *A shorter model theory*. Cambridge: Cambridge University Press.

Indrzejczak, A. (1997). Generalised sequent calculus for propositional modal logics. *Logica Trianguli, 1*, 15–31.

Indrzejczak, A. (2016). Linear time in hypersequent framework. *Bulletin of Symbolic Logic, 22*, 121–144.
Kanger, S. (1957). *Provability in logic*. Almqvist and Wiksell.
Kashima, R. (1994). Cut-free sequent calculi for some tense logics. *Studia Logica, 53*, 119–135.
Kleene, S. C. (1952). *Introduction to metamathematics*. Amsterdam: North-Holland.
Kracht, M. (1996). Power and weakness of the modal display calculus. In H. Wansing (Ed.), *Proof theory of modal logic* (pp. 93–121). Kluwer.
Kripke, S. A. (1963a). Semantical analysis of modal logic I: Normal modal propositional calculi. *Zeitschrift für mathematische Logik und Grundlagen der Mathematik, 11*, 3–16.
Kripke, S. A. (1963b). Semantical considerations on modal logic. *Acta Philosophica Fennica, 16*, 83–94.
Kripke, S. A. (1965). Semantical analysis of modal logic II. non-normal modal propositional calculi. In J. Addison, L. Henkin, & A. Tarski (Eds.), *The theory of models, proceedings of the 1963 international symposium at berkeley* (pp. 206–220). North-Holland Publ. Co.
Kurokawa, H. (2014). Hypersequent calculi for modal logics extending S4. In Y. Nakano, K. Satoh, & D. Bekki (Eds.), *New frontiers in artificial intelligence. JSAI-isAI 2013. Lecture notes in computer science* (Vol. 8417, pp. 51–68). Cham: Springer. https://doi.org/10.1007/978-3-319-10061-6_4
Kuznets, R. (2018). Multicomponent proof-theoretic method for proving interpolation properties. *Annals of Pure and Applied Logic, 169*, 1369–1418.
Kuznets, R., & Lellmann, B. (2016). Grafting hypersequents onto nested sequents. *Logic Journal of the IGPL, 24*, 375–423.
Kuznets, R., & Straßburger, L. (2018). Maehara-style modal nested calculi. *Archive for Mathematical Logic*. https://doi.org/10.1007/s00153-018-0636-1
Lahav, O., & Avron, A. (2013). A unified semantic framework for fully structural propositional sequent systems. *ACM Transactions on Computational Logic*, *14*, 1–33.
Leivant, D. (1981). On the proof theory of the modal logic for arithmetic provability. *The Journal of Symbolic Logic, 46*(3), 531–538.
Lellmann, B. (2015). Linear nested sequents, 2-sequents and hypersequents. In H. De Nivelle (Ed.), *Automated reasoning with analytic tableaux and related methods. TABLEAUX 2015. Lecture notes in computer science* (Vol. 9323, pp. 135–150). Berlin: Springer. https://doi.org/10.1007/978-3-319-24312-2_10
Lellmann, B. (2016). Hypersequent rules with restricted contexts for propositional modal logics. *Theoretical Computer Science, 656*, 76–105.
Lellmann, B. (2019). Combining monotone and normal modal logic in nested sequents – with countermodels. In S. Cerrito, A. Popescu (Eds.), *Automated reasoning with analytic tableaux and related methods. TABLEAUX 2019. Lecture notes in computer science* (Vol. 11714, pp. 203–220). Berlin: Springer. https://doi.org/10.1007/978-3-030-29026-9_12
Lellmann, B., & Pimentel, E. (2019). Modularisation of sequent calculi for normal and non-normal modalities. *ACM Transactions on Computational Logic*, *20*(2), 1–46.
Lellmann, B., Pimentel, E., & Olarte, C. (2017). A uniform framework for substructural logics with modalities. In *LPAR-21. 21st international conference on logic for programming, artificial intelligence and reasoning. EPiC series in computing* (Vol. 46, pp. 435–455).
Lyon, T. (2020a). On deriving nested calculi for intuitionistic logics from semantic systems. In S. Artëmov & A. Nerode (Eds.), *Logical foundations of computer science. LFCS 2020. Lecture notes in computer science* (Vol. 11972, pp. 177–194). Berlin: Springer. https://doi.org/10.1007/978-3-030-36755-8_12
Lyon, T. (2020b). On the correspondence between nested calculi and semantic systems for intuitionistic logics. *Journal of Logic and Computation*. https://doi.org/10.1093/logcom/exaa078
Lyon, T., Tiu, A., Goré, R., & Clouston, R. (2020). Syntactic interpolation for tense logics and bi-intuitionistic logic via nested sequents. In M. Fernández & A. Muscholl (Eds.), *28th EACSL Annual Conference on Computer Science Logic, CSL 2020, January 13–16, 2020, Barcelona,*

*Spain* (Vol. 152, pp. 28:1–28:16). Schloss Dagstuhl - Leibniz-Zentrum für Informatik. https://doi.org/10.4230/LIPIcs.CSL.2020.28

Lyon, T. S. (2021a). Nested sequents for intuitionistic modal logics via structural refinement. In A. Das & S. Negri (Eds.), *Automated reasoning with analytic tableaux and related methods. TABLEAUX 2021. Lecture notes in computer science* (Vol. 12842, pp. 409–427). Berlin: Springer. https://doi.org/10.1007/978-3-030-86059-2_24

Lyon, T. S. (2021b). *Refining labelled systems for modal and constructive logics with applications*. Doctoral dissertation, Technische Universität Wien.

Maehara, S. (1961). On the interpolation theorem of Craig. *Sũgaku, 12*, 235–237.

Marin, S., Miller, D., & Volpe, M. (2016). A focused framework for emulating modal proof systems. In L. D. Beklemishev, S. Demri, & A. Maté (Eds.), *Advances in Modal Logic 11, Proceedings of the 11th Conference on "Advances in Modal Logic," Held in Budapest, Hungary, August 30 - September 2, 2016* (pp. 469–488). College Publications. http://www.aiml.net/volumes/volume11/Marin-Miller-Volpe.pdf

Marin, S., & Straßburger, L. (2014). Label-free modular systems for classical and intuitionistic modal logics. In R. Goré, B. P. Kooi, & A. Kurucz (Eds.), *AiML 10* (pp. 387–406). College.

Marin, S., & Straßburger, L. (2017). Proof theory for indexed nested sequents. In R. A. Schmidt & C. Nalon (Eds.), *Automated reasoning with analytic tableaux and related methods. TABLEAUX 2017. Lecture notes in computer science* (Vol. 10501, pp. 81–97). Berlin: Springer. https://doi.org/10.1007/978-3-319-66902-1_5

Masini, A. (1992). 2-sequent calculus: A proof theory of modalities. *Annals of Pure and Applied Logic, 58*, 229–246.

Massacci, F. (2000). Single step tableaux for modal logics. *Journal of Automated Reasoning, 24*, 319–364.

Miller, D., & Volpe, M. (2015). Focused labeled proof systems for modal logic. In M. Davis, A. Fehnker, A. McIver, & A. Voronkov (Eds.), *Logic for programming, artificial intelligence, and reasoning. LPAR 2015. Lecture notes in computer science* (Vol. 9450, pp. 266–280). Berlin: Springer. https://doi.org/10.1007/978-3-662-48899-7_19

Mints, G. (1971). On some calculi of modal logic. *Proceedings of the Steklov Institute of Mathematics, 98*, 97–122.

Mints, G. (1997). Indexed systems of sequents and cut-elimination. *Journal of Philosophical Logic, 26*(6), 671–696.

Mints, G., Olkhovikov, G., & Urquhart, A. (2013). Failure of interpolation in constant domain intuitionistic logic. *The Journal of Symbolic Logic, 78*, 937–950.

Negri, S. (2005). Proof analysis in modal logic. *Journal of Philosophical Logic, 34*, 507–544.

Ohnishi, M., & Matsumoto, K. (1957). Gentzen method in modal calculi. *Osaka Journal of Mathematics, 9*, 113–130.

Parisi, A. (2017). *Second-order modal logic*. Doctoral dissertation, University of Connecticut. Retrieved from http://digitalcommons.uconn.edu/dissertations/1480

Parisi, A. (2020). A hypersequent solution to the inferentialist problem of modality. *Erkenntnis*. (Published online: 05 June 2020.) https://doi.org/10.1007/s10670-020-00264-x

Pimentel, E., Ramanayake, R., & Lellmann, B. (2019). Sequentialising nested sequents. In S. Cerrito & A. Popescu (Eds.), *Automated reasoning with analytic tableaux and related methods. TABLEAUX 2019. Lecture notes in computer science* (Vol. 11714, pp. 147–165). Berlin: Springer. https://doi.org/10.1007/978-3-030-29026-9_9

Poggiolesi, F. (2008a). A cut-free simple sequent calculus for modal logic S5. *Review of Symbolic Logic, 1*, 3–15.

Poggiolesi, F. (2008b). The method of tree-hypersequent for modal propositional logic. In D. Makinson, J. Malinowski, & H. Wansing (Eds.), *Towards mathematical philosophy. Trends in logic (TREN)* (Vol. 28, pp. 31–51). Dordrecht: Springer. https://doi.org/10.1007/978-1-4020-9084-4_3

Poggiolesi, F. (2008c). A new definition of the Dośen's principle. In M. Pelis (Ed.), *Logica yearbook 2007* (pp. 133–142). Prague: Filosofia.

Poggiolesi, F. (2008d). *Sequent calculi for modal logic*. Doctoral dissertation, University of Florence and University of Paris 1.
Poggiolesi, F. (2009). A purely syntactic and cut-free sequent calculus for the modal logic of provability. *Review of Symbolic Logic, 2*, 593–611.
Poggiolesi, F. (2010a). Display calculi and other modal calculi: A comparison. *Synthese, 173*, 259–279.
Poggiolesi, F. (2010b). *Gentzen calculi for modal propositional logic. Trends in logic (TREN)* (Vol. 32). Dordrecht: Springer. https://doi.org/10.1007/978-90-481-9670-8
Poggiolesi, F. (2010c). Reflecting the semantic features of S5 at the syntactic level. In C. Sinigaglia & M. D'Agostino (Eds.), *Proceedings of SILFS Conference* (pp. 13–25). London: College Publications.
Poggiolesi, F. (2013). From a single agent to multi-agent via hypersequents. *Logica Universalis, 7*, 147–166.
Postniece, L. (2009). Deep inference in bi-intuitionistic logic. In H. Ono, M. Manazawa, & R. de Queiroz (Eds.), *Logic, language, information and computation. WoLLIC 2009. Lecture notes in computer science* (Vol. 5514, pp. 320–334). Berlin: Springer. https://doi.org/10.1007/978-3-642-02261-6_26
Postniece, L. (2010). *Proof theory and proof search of bi-intuitionistic and tense logic*. Doctoral dissertation, The Australian National University.
Pottinger, G. (1983). Uniform cut-free formulations of T, S4 and S5 (abstract). *Journal of Symbolic Logic, 48*, 900–901.
Ramanayake, R. (2018). Inducing syntactic cut elimination for indexed nested sequents. *Logical Methods in Computer Science*, *14*(4:18), 1–25.
Restall, G. (2007). Proofnets for S5: sequents and circuits for modal logic. In *Logic colloquium 2005* (Vol. 28, pp. 151–172). Cambridge University Press.
Sambin, G., & Valentini, S. (1982). The modal logic of provability. the sequential approach. *Journal of Philosophical Logic*, *11*(3), 311–342. https://doi.org/10.1007/BF00293433
Segerberg, K. (1968). Decidability of four modal logics. *Theoria, 34*, 21–25.
Shamkanov, D. (2015). Nested sequents for provability logic GLP. *Logic Journal of the IGPL, 23*(5), 789–815.
Simpson, A. K. (1994). *The proof theory and semantics of intuitionistic modal logic*. Doctoral dissertation, Edinburgh.
Smullyan, R. M. (1968). *First-order logic. Ergebnisse der Mathematik und ihrer Grenzgebiete. 2. Folge (MATHE2)* (Vol. 43). Berlin: Springer. https://doi.org/10.1007/978-3-642-86718-7
Stewart, C., & Stouppa, P. (2004). A systematic proof theory for several modal logics. In R. Schmidt, I. Pratt-Hartmann, M. Reynolds, & H. Wansing (Eds.), *Advances in modal logic* (Vol. 5, pp. 309–333). London: College Publications.
Stouppa, P. (2007). A deep inference system for the modal logic S5. *Studia Logica, 85*, 199–214.
Straßburger, L. (2013). Cut elimination in nested sequents for intuitionistic modal logics. In F. Pfenning (Ed.), *Foundations of software science and computation structures. FoSSaCS 2013. Lecture notes in computer science* (Vol. 7794, pp. 209–224). Berlin: Springer. https://doi.org/10.1007/978-3-642-37075-5_14
Tanaka, Y. (2003). Some proof systems for predicate common knowledge. *Reports on Mathematical Logic, 37*, 79–100.
Tiu, A., Ianovski, E., & Goré, R. (2012). Grammar logics in nested sequent calculus: Proof theory and decision procedures. In T. Bolander, T. Braüner, S. Ghilardi, & L. S. Moss (Eds.), *Advances in modal logic 9, papers from the ninth conference on "advances in modal logic," held in copenhagen, denmark, 22–25 august 2012* (pp. 516–537). College Publications.
Troelestra, A. S., & Schwichtenberg, H. (1996). *Basic proof theory*. Cambridge: Cambridge University Press.
Viganò, L. (2000). *Labelled non-classical logics*. Kluwer Academic Publishers.
Wansing, H. (1998). *Displaying modal logic. Trends in logic (TREN)* (Vol. 3). Berlin: Springer. https://doi.org/10.1007/978-94-017-1280-4

Wansing, H. (2002). Sequent systems for modal logics. In D. M. Gabbay & F. Guenthner (Eds.), *Handbook of philosophical logic (HALO)* (Vol. 8). Dordrecht: Springer. https://doi.org/10.1007/978-94-010-0387-2_2

# Chapter 12
# A Proof-Theoretic Approach to Formal Epistemology

**Sara Negri and Edi Pavlović**

**Abstract** Ever since antiquity, attempts have been made at defining knowledge through belief augmented by additional properties such as truth and justification. These characterizations have been challenged by Gettier counterexamples and their variants. A modern proposal, what is known as defeasibility theory, characterizes knowledge through stability under revision of beliefs on the basis of true or arbitrary information. A formal investigation of such a proposal calls for the methods of dynamic epistemic logic: well developed semantic approaches to dynamic epistemic logic have been given through the plausibility models, but a corresponding proof theory is still in its infancy. We shall recast plausibility models in terms of more general neighbourhood models and develop on their basis complete proof systems. An inferential treatment of epistemic and doxastic notions such as safe belief will give a new way to study their relationships; among these, the characterization of knowledge as belief stable under arbitrary revision will be grounded through formal labelled sequent calculus derivations.

**Keywords** Conditional doxastic logic · Epistemic paradoxes · Formal epistemology · Gettier problem · Labelled sequent calculus

S. Negri
Department of Mathematics, University of Genova, via Dodecaneso 35, 16100 Genova, Italy
e-mail: sara.negri@unige.it

E. Pavlović (✉)
Munich Center for Mathematical Philosophy (MCMP), Ludwig-Maximilians-Universität München, Ludwigstraße 31, 80539 München, Germany
e-mail: edi.pavlovic@lmu.de

© The Author(s), under exclusive license to Springer Nature Switzerland AG 2024

Y. Weiss and R. Birman (eds.), *Saul Kripke on Modal Logic*, Outstanding Contributions to Logic 30, https://doi.org/10.1007/978-3-031-57635-5_12

## 12.1 Introduction

We start with a (the) problem in formal epistemology: how can it define its domain of inquiry? What is knowledge? Knowledge is truthful, that is an unquestioned assumption, and we believe what we know. So knowledge is at least true belief. But there should be something more. Ever since antiquity, attempts have been made at defining knowledge through belief augmented by additional properties such as truth and justification, starting with Plato's dialogue *Meno* ($\sim$ 400 BC):

> *Socrates*: [...] Now this is an illustration of the nature of true opinions: while they stay with us they are beautiful and fruitful, but they run away out of the human soul, and do not remain long, and therefore they are not of much value until they are fastened by the tie of the cause; and this fastening of them, friend Meno, is recollection, as you and I have agreed to call it. But when they are bound, in the first place, they have the nature of knowledge; and, in the second place, they are permanent. And this is why knowledge is more honourable and excellent than true opinion, because fastened by a chain.

and with the *Theaetetus* ($\sim$ 369 BC):

> *Socrates*: When, therefore, any one forms the true opinion of anything without rational explanation, you may say that his mind is truly exercised, but has no knowledge; for he who cannot give and receive a reason for a thing, has no knowledge of that thing; but when he adds rational explanation, then, he is perfected in knowledge.[1]

In the former, knowledge is defined positively as a true belief augmented by a justification, and in the latter, it is defined in the negative: if there is no rational explanation for true beliefs, there is no knowledge. Together, the positive and negative definition give a characterization of knowledge.

Indeed, the problem of defining knowledge can be considered as the central question of formal epistemology and the most common answer to this question is given by the following equation:

**Knowledge = justified true belief**

However, as is often the case in formal epistemology, whenever a characterization is proposed it gets challenged by ad hoc arguments. In a famous, much cited 2-page article (Gettier, 1963) Gettier proposes counterexamples to the equation. Gettier's first counterexample is as follows: Smith and Jones have applied for a job. Smith has strong evidence for believing the following conjunctive proposition:

$A$: Jones is the man who will get the job, and Jones has ten coins in his pocket.

Moreover, Smith correctly deduces that proposition $A$ entails, and thus accepts:

$B$: The man who will get the job has ten coins in his pocket.

Smith therefore has justified belief in $B$. Proposition $A$ turns out to be false because Smith, and not Jones, gets the job. However, Smith, not knowing that, also has ten coins in his pocket. Proposition $B$ is then true, Smith believes $B$, and Smith is justified in believing $B$, but doesn't know $B$.

[1] Both excerpts are from Benjamin Jowett's translation of Plato's dialogues (Jowett & Pelliccia, 2001).

The structure of the counterexamples is as follows: there are two facts $A$ and $B$ that are connected by a justification chain, so that belief in $A$ entails belief in $B$. Something happens that breaks the justification chain because $A$ turns out to be false while $B$ remains true, so that $B$ is true but not because of $A$. So an agent can have a justified true belief in a proposition $B$ while not knowing it. So a related, more general, question is whether one is justified in believing something that is only by chance true. Is this knowledge?

Some 50 years earlier a very similar set of examples, and largely to the same effect, was offered by Bertrand Russell, where in (Russell, 1912, Chap. 13) he writes[2]:

> If a man believes that the late Prime Minister's last name began with a B, he believes what is true, since the late Prime Minister was Sir Henry Campbell Bannerman. But if he believes that Mr. Balfour was the late Prime Minister, he will still believe that the late Prime Minister's last name began with a B, yet this belief, though true, would not be thought to constitute knowledge. If a newspaper, by an intelligent anticipation, announces the result of a battle before any telegram giving the result has been received, it may by good fortune announce what afterwards turns out to be the right result, and it may produce belief in some of its less experienced readers. But in spite of the truth of their belief, they cannot be said to have knowledge. Thus it is clear that a true belief is not knowledge when it is deduced from a false belief.

In these cases we are again faced with a belief which is true by mere happenstance, and thus falls short of knowledge.

Epistemic defeasibility (Lehrer & Paxson 1969; Sosa, 1964; Swain, 1974) was introduced into epistemology as an ingredient in one of the main strategies for dealing with Gettier-style cases. In these cases, an individual's justified true belief fails to count as knowledge because the justification is defective as a source of knowledge. According to the defeasibility theory of knowledge, the defect involved can be characterized in terms of evidence that the subject does not possess which overrides, or defeats, the subject's prima facie justification for belief. This account holds that knowledge is *indefeasibly justified true belief* and characterizes knowledge through stability under revision of beliefs on the basis of true or arbitrary information (Hintikka, 1962; Stalnaker, 1998).

The basic question that should be answered then becomes, "Would we still believe in $A$ if we were given some additional evidence?" This is, technically speaking, a counterfactual with arbitrary assumptions, and the procedure to answer the question can be described in terms of Lehrer's "justification game" (Lehrer, 2018, p. 119).

> The justification game is played in the following way. The claimant presents something she accepts as true. The skeptic may then raise any objection in the form of a competitor of what the claimant presents. If what the claimant accepts is something that is more reasonable for her to accept than the skeptical objection, that is, if the competitor cited by the skeptic is beaten, then the claimant wins the round. If all the competitors raised by the skeptic are beaten, then the claimant wins the game. If she wins the game, she is personally justified in accepting what she presented; if not, she is not personally justified. The game is a heuristic device for understanding the considerations that make a person justified in accepting something rather than a psychological model of mental processes.

[2] We thank Saul Kripke and Allen Hazen for bringing this passage to our attention.

Clearly, a formal account of the new characterization first requires an explanation of what it means to revise a belief on the face of new information. One of the most popular logical approches to belief revision is the theory AGM (Alchourrón et al., 1985). An operation of belief revision takes an epistemic state $K$ (i.e. what an agent believes at a certain time) and a proposition $A$ and gives a new, revised, epistemic state $K * A$; this defines a consequence relation $A \mid\sim_K B \equiv B \in K * A$ with the following features: the relation is a nonmonotonic consequence relation; it depends on a background epistemic state, so it is indeed a family of consequence relations, as indicated by the subscript $K$; the relation only concerns the "facts" of the world and not higher-order beliefs (i.e. beliefs about beliefs); the conditional is at the level of the consequence relation, not at the level of the object language, so it cannot treat nested conditionals nor iterated belief revision.

For these reasons something more general is needed, capable of capturing both epistemic and doxastic modalities, a conditional in the object language, a notion of justification and of revision, tools for performing valid inferences and for capturing both an absolute (impersonal) and a multi-agent setting. Moreover, it needs to feature both a semantic and a syntactic approach, related to each other by completeness theorems and finally a notion of computation, to handle complex reasoning patterns.

All this can be achieved using a suitable proof-theoretic approach to *dynamic epistemic logic*. The latter is indeed not a single logic but "the study of a family of modal logics, each of which is obtained from a given logical language by adding one or more modal operators that describe model-transforming actions," as compactly defined in the survey offered in the SEP entry.[3]

The semantic approach to dynamic epistemic logic (preferred in the literature) utilizes tools which are generalizations of possible world semantics and include what are known as *plausibility models* (Baltag & Smets, 2008; Board, 2004; Pacuit, 2013), *Grove models* (Grove, 1988), and *Lewis' sphere semantics* (Lewis, 1973). In recent years a more general semantic approach, based on *neighbourhood models* (Baltag & Smets, 2008; Pacuit, 2013) has gained prominence. In contrast to a well-established semantic apparatus, a corresponding proof theory for the realm of dynamic epistemic logic is still in its infancy, but already features a variety of approaches based on labelled sequent calculi (Girlando et al., 2018b; Nomura et al., 2020), nested sequent calculi (Dyckhoff & Sadrzadeh, 2018a; Girlando et al., 2010), labelled tableaux (Aucher et al., 2012), and display calculi (Frittella et al., 2016).[4]

[3] Cf. Baltag and Renne (2016) at: https://plato.stanford.edu/entries/dynamic-epistemic/.

[4] This paper is not the place for a history of labelled proof systems, so we shall just limit ourselves to mentioning that the internalization of possible worlds semantics in inferential calculi, an idea dating back to Kanger (1957), has been developed in various forms: tableaux systems (Catach, 1991; Fitting, 1983; Goré, 1999; Massacci, 2000; Nerode, 1991; Orlowska & Golińska Pilarek, 2011; Priest, 2008), systems of natural deduction (Basin et al., 1998; Fitch, 1966; Simpson, 1994), sequent systems (Castellini, 2006; Castellini & Smaill, 2002; Kushida & Okada, 2003; Mints, 1997; Viganó, 2000), and in what is known as *Hybrid Logic* (Blackburn, 2000). *Labelled Deductive Systems* (Gabbay, 1996; Russo, 1995) instead mainly use the internalization of algebraic semantics natural deduction style. Labelled deduction has also been used for systems of automated deduction

We shall focus on a particular dynamic epistemic logic, the logic of conditional belief (CDL, for Conditional Doxastic Logic) (Baltag & Smets, 2006, 2008; Board, 2004; Girlando et al., 2018b; Pacuit, 2013), as a formal way of facing the question of defining knowledge. The logic has as its main modality that of conditional belief relative to an agent, so it contains belief revision at the level of the object language. Its axiomatization will be recalled, as well as its semantics in terms of neighbourhood models, and the relation of the latter to Kripke models. We shall recast plausibility models in terms of the more general neighbourhood models and develop on their basis complete proof systems following a methodology introduced by Negri (2017b) and developed for conditional doxastic notions by Girlando et al. (2018b). The result is a *labelled* sequent calculus: in contrast to other formalisms such as nested sequents, hypersequents or display calculi, the usual structural syntax of sequents is not extended, but instead formulas that appear in sequents contain formal expressions for the *internalized* semantics, i.e. a semantics that becomes an inherent part of the syntax of the calculus. In the case of a labelled sequent calculus that internalizes neightbourhood semantics, both possible worlds and neighbourhoods appear as labels; these are used to express forcing relations: a pointwise one (a world forces a formula) and two local ones (a neighbourhood forces completely or at some point a formula). The rules of the conditional modality are then obtained by unfolding the semantic explanation in terms of such forcing relations. Additional rules are used to express properties of the models. Knowledge and belief are obtained as degenerate cases of conditional belief, and the rules are found accordingly.

Once the rules of the calculus have been given we establish, in turn and with each subsequent proof utilizing the previous ones, the typical structural properties—admissibility of weakening, invertibility of rules, admissibility of contraction and cut. The latter in particular is paramount in proof theory, and its applications will extend beyond the confines of this paper. Labelled systems, in fact, have several advantages over other formalisms for modal logic. The systems originate from a uniform methodology that, together with the standard syntax of sequents, makes the proof of said structural properties of the calculi a rather routine task, once the rules have been set up in the right way. Because of the far reaching relational and neighbourhood semantics, the method has shown its versatility for a wide range of applications; the transparent semantic motivation behind the rules makes them intuitive and allows a direct completeness proof in the Tait-Schütte-Takeuti style: for any sequent in the language, we can either find a derivation in the calculus or else, if the proof search fails, a countermodel in the appropriate class of neighbourhood models. Because of invertibility of the rules, a countermodel to a premiss of a rule is a countermodel to the conclusion. One can thus obtain a countermodel from any branch of the proof-search tree that does not end with an initial sequent, a purely local procedure. When a finite bound can be established to the procedure, such a completeness proof becomes a proof of decidability, as in (Girlando et al., 2018b).

for non-classical logics, such as the system LoTREC (del Cerro et al., 2001). A comparison of various formalisms for modal logic, including labelled ones, is presented by Indrzejczak (2010).

After all the basic machinery has been set up, we will finally be in a position to give the rules of an operator of *safe belief* and can prove, using just the rules of the calculus, that safe belief can indeed be characterized as the doxastic attitude which is stable under revision with arbitrary true information. Knowledge as defined in CDL instead is seen to be stable under revision by arbitrary (not just true) information.

Finally, we show how proof theory gives a powerful tool for paradox control. In fact, the use of an analytic proof system, in which the formulas appearing in a proof are constrained by what occurs in a theorem under consideration, gives the possibility to reason not just on derivability but also on underivability since by the use of a ground sequent calculus with invertible rules we get both (a) a way to find the required rules to obtain specific systems and (b) formal proofs of underivability of certain properties, thus giving definite arguments for paradox defeat. One such example is the argument against the paradox of the perfect believer (the tricky implication from belief of knowledge to knowledge) and the 'I know only what I believe'-paradox infamously known from world's politicians' self-centered declarations. Both will be discussed in Sect. 12.8.

## 12.2 Neighbourhood Semantics

We begin our formal presentation by introducing the semantic framework in which all of our following investigations will be situated, that of neighbourhood semantics. Discovered independently in 1970 by Scott (1970) and Montague (1970) in an effort to give an interpretation of non-normal modal logics, neighbourhood models assign to each world a set of sets of worlds, and are used to interpret modalities. It is well known that neighbourhood semantics is a generalization of Kripke (or relational) semantics. Below, we shall detail this result by recalling how Kripke frames correspond to a certain subset of neighbourhood frames.[5] Let us start with the definition of neighbourhood frames.

**Definition 12.2.1** A *neighbourhood frame* has the form $\langle W, I\rangle$ where $W$ is a nonempty set and $I : W \to \mathcal{P}(\mathcal{P}(W))$ is a neighbourhood function.

Given a neighbourhood frame, we form a neighbourhood model $\mathcal{M} = \langle W, I, [\![\ ]\!]\rangle$ by adding the propositional evaluation function $[\![\ ]\!]$:

**Definition 12.2.2** $[\![\ ]\!] : Atm \to \mathcal{P}(W)$ is the evaluation for atomic formulas. Truth conditions for formulas extend $[\![\ ]\!]$ inductively as:

[5] For a recent systematic account of neighbourhood semantics, especially in relation to its use in modal logic, see Pacuit (2017).

$$\begin{aligned}
[\![\neg A]\!] &= W - [\![A]\!] \\
[\![A \wedge B]\!] &= [\![A]\!] \cap [\![B]\!] \\
[\![A \vee B]\!] &= [\![A]\!] \cup [\![B]\!] \\
[\![A \supset B]\!] &= (W - [\![A]\!]) \cup [\![B]\!]
\end{aligned}$$

A formulas $A$ is valid in $\mathcal{M}$ iff $[\![A]\!] = W$. We write $x \in [\![A]\!]$ as $\mathcal{M}, x \Vdash A$, and further omit $\mathcal{M}$ if no ambiguity arises.

### 12.2.1 Neighbourhood and Kripke Frames

We shall now detail the correspondence between Kripke frames and *augmented* neighbourhood frames (Chellas, 1980).

**Definition 12.2.3** (*Augmented neighbourhood frame*) A neighbourhood frame is augmented iff for every neighbourhood $a$ and point $x$,

$$a \in I(x) \equiv \bigcap I(x) \subseteq a$$

To represent a Kripke frame $\langle W, R\rangle$ in a neighbourhood frame (Negri, 2017a), one defines the neighbourhood function $I^R$ as:

**Definition 12.2.4** (*Neighbourhood function* $I^R$)

$$I^R(x) = \{a \mid R(x) \subseteq a\}$$

This means that we take all supersets of worlds accessible from $x$ as its neighbourhoods. Likewise, one can represent a neighbourhood frame $\langle W, I\rangle$ in a relational one by defining the accessibility relation $R^I$ as:

**Definition 12.2.5** (*Accessibility relation* $R^I$) $x R^I y \equiv y \in \bigcap I(x)$.

The accessibility relation inherited from a neighbourhood frame is defined by stating that $y$ is accessible from $x$ iff it is in all of its neighbourhoods.

We are now ready to recall the correspondence between Kripke frames and augmented neighbourhood frames. The proof of Lemmas 12.2.6 and 12.2.7 and of Theorem 12.2.8 below are already known, e.g. from Chellas (1980), and their proofs are included for the convenience of the reader.

**Lemma 12.2.6** *For every Kripke model* $\mathcal{M} = \langle W, R, \mathcal{V}\rangle$ *there is an augmented neighbourhood model* $\mathcal{M}^R = \langle W, I^R, [\![\ ]\!]\rangle$ *such that for any* $w$, *if* $\mathcal{M}, w \Vdash A$, *then* $\mathcal{M}^R, w \Vdash A$.

***Proof*** First, by Definition 12.2.4, $a \in I^R(x) \equiv R(x) \subseteq a$. Then, since $\bigcap I^R(x) = R(x)$, it follows $\mathcal{M}^R$ is augmented.

The lemma is proven by induction on the weight of $A$. We illustrate just the interesting case of $\Box$, where

$$M^R, w \Vdash \Box A \equiv [\![A]\!] \in I(w)$$

If $\mathcal{M}, w \Vdash \Box A$, then $\forall y(wRy \supset \mathcal{M}, y \Vdash A)$. So, by the inductive hypothesis, $R(w) \subseteq [\![A]\!]$. Therefore, by Definition 12.2.4, $[\![A]\!] \in I^R(w)$, and finally $\mathcal{M}^R, w \Vdash \Box A$. □

**Lemma 12.2.7** *For every augmented neighbourhood model $\mathcal{M} = \langle W, I, [\![\ ]\!] \rangle$ there is a Kripke model $\mathcal{M}^I = \langle W, R^I, \mathcal{V} \rangle$ such that for any $w$, if $\mathcal{M}, w \Vdash A$, then $\mathcal{M}^I, w \Vdash A$.*

***Proof*** By induction on the weight of $A$. We again illustrate only the $\Box$ case.

If $\mathcal{M}, w \Vdash \Box A$ then $[\![A]\!] \in I(w)$. Then by Definition 12.2.3, $\bigcap I(w) \subseteq [\![A]\!]$. So, by Definition 12.2.5 and the inductive hypothesis, $\forall y(wR^I y \supset \mathcal{M}^I, y \Vdash A)$. Therefore $\mathcal{M}^I, w \Vdash \Box A$. □

Combined, these lemmas show that:

**Theorem 12.2.8** *For every Kripke model, there is an augmented neighbourhood model that validates the same formulas, and vice versa.*

## 12.3 Conditional Doxastic Logic CDL

The dynamic epistemic logic we will be utilizing to investigate the question of defining knowledge is Conditional Doxastic Logic (CDL, for short) (Board, 2004; Baltag & Smets, 2006, 2008; Pacuit, 2013); we build here on the extensive proof-theoretic treatment developed by Girlando et al. (2018b).

The logic has as a main modality conditional belief relative to an agent, so it contains belief revision at the level of the object language. The primitive epistemic operator of conditional belief $Bel_i(C|B)$ is understood as "*agent i believes C, given B*". In terms of belief revision, this would correspond to having the proposition $C$ in the starting epistemic set $K$ revised with $B$, written, in the notation of belief revision, as $B * K$. We first recall the definition of its language.

**Definition 12.3.1** (*Language of CDL*)

$$A \ := \ P \mid \bot \mid \neg A \mid A \wedge A \mid A \vee A \mid A \supset A \mid Bel_i(A|A)$$

An axiomatization of CDL that augments the axioms of epistemic logic by those of AGM belief revision (Alchourrón et al., 1985) can be found in (Board, 2004).

**Definition 12.3.2** (*Inference rules of CDL*)

1. If $\vdash B$, then $\vdash Bel_i(B|A)$ (epistemization rule)
2. If $\vdash A \supset\subset B$, then $\vdash Bel_i(C|A) \supset\subset Bel_i(C|B)$[6] (rule of logical equivalence)

The epistemization rule is a form of necessitation: if $B$ is provable, the agent believes it under any assumption. The second rule corresponds to K*6 of AGM belief revision and states that logically equivalent formulas should lead to identical belief revisions, i.e. only the content, not the form, of information matters.

**Definition 12.3.3** (*Axioms of CDL*) Contain any axiomatization of the classical propositional calculus, plus:

3. $(Bel_i(B|A) \wedge Bel_i(B \supset C|A)) \supset Bel_i(C|A)$ (distribution axiom)
4. $Bel_i(A|A)$ (success axiom)
5. $Bel_i(B|A) \supset (Bel_i(C|A \wedge B) \supset\subset Bel_i(C|A))$ (minimal change principle 1)
6. $\neg Bel_i(\neg B|A) \supset (Bel_i(C|A \wedge B) \supset\subset Bel_i(B \supset C|A))$ (minimal change principle 2)
7. $Bel_i(B|A) \supset Bel_i(Bel_i(B|A)|C)$ (positive introspection)
8. $\neg Bel_i(B|A) \supset Bel_i(\neg Bel_i(B|A)|C)$ (negative introspection)
9. $A \supset \neg Bel_i(\bot|A)$ (consistency axiom)

Axiom 3 expresses closure of conditional belief under modus ponens; the success axiom 4 corresponds to AGM K*2 and states that the information received is indeed believed in the revised information state. Board (2004) calls 5 and 6 axioms of informational economy; when we change beliefs in response to new evidence the change should be minimal, not greater than what is needed to incorporate the new evidence; 5 states that if we learn something that we already believed (B), we don't have to revise our beliefs at all. Axiom 6 corresponds to AGM K*7 and K*8: If and agent learns something consistent with her beliefs (B), the revised beliefs obey the deduction theorem. Finally the names of 7, 8, 9 are suggestive of their epistemic meaning.

The semantics of CDL can be formulated in terms of *epistemic plausibility models* (Baltag & Smets, 2008; Board, 2004; Pacuit, 2013). The main idea behind the semantics, with two sets of agent-indexed relations, is that the equivalence relations $\sim_i$ give a partition of the set of possible worlds according to agent $i$'s 'hard' information, whereas the plausibility orderings $\preceq_i$ are related to 'soft' information and tell which worlds inside a class of epistemically equivalent worlds the agent $i$ considers the most plausible.

**Definition 12.3.4** (*Epistemic plausibility models*) Let $\mathcal{A}$ be a set of agents. An *epistemic plausibility model* ($EPM$) has the form

$$\mathcal{M} = \langle W, \{\sim_i\}_{i\in\mathcal{A}}, \{\preceq_i\}_{i\in\mathcal{A}}, [\![\ ]\!]\rangle$$

where $W$ is a non-empty set of elements called "worlds";

[6] Where $A \supset\subset B$ abbreviates $(A \supset B) \wedge (B \supset A)$.

for each $i \in \mathcal{A}$, $\sim_i$ is an equivalence relation over $W$;
for each $i \in \mathcal{A}$, $\preceq_i$ is a well-founded pre-order over $W$;
$[\![\ ]\!] : Atm \to \mathcal{P}(W)$ is the evaluation for atomic formulas.
The relations $\sim_i$ and $\preceq_i$ satisfy the following properties:

1. **Plausibility implies possibility**: If $w \preceq_i v$ then $w \sim_i v$ (preference ordering makes sense only within cells consisting of epistemically equivalent states).
2. **Local connectedness**: If $w \sim_i v$ then $w \preceq_i v$ or $v \preceq_i w$ (in other words, $\preceq_i$ is a linear order over every equivalence class of $\sim_i$).

**Definition 12.3.5** Truth conditions for formulas in EPM are just like those in Definition 12.2.2, with the addition of

$$[\![Bel_i(B|A)]\!] \equiv \{x \in W \mid Min_{\preceq_i}([x]_{\sim_i} \cap [\![A]\!]) \subseteq [\![B]\!]\}$$

where $[x]_{\sim_i} = \{w \mid w \sim_i x\}$, and

$$Min_{\preceq_i}(S) = \{u \in S \mid \forall z \in S\ (u \preceq_i z)\}$$

The above semantic clause has the following meaning: agent $i$ believes $B$ conditionally on $A$ in a world $x$ if (and only if) $B$ holds in *the most plausible worlds* accessible from $x$ in which $A$ holds, the "most plausible worlds" for an agent $i$ being the $\preceq_i$-minimal ones.

The connection of the two, which we will recall later, has been established by Board (2004), where it is shown that a formula $A$ is a theorem of *CDL* if and only if it is valid in the class of epistemic plausibility models.

### *12.3.1 Neighbourhood Models of CDL*

We now combine the terminology established in the current and the previous section to introduce the semantic framework of the remainder of this paper, of (multi-agent) neighbourhood models of CDL.

**Definition 12.3.6** (*Multi-agent neighbourhood models*) Let $\mathcal{A}$ be a set of agents; a *multi-agent neighbourhood model* ($NM$) has the form

$$\mathcal{M} = \langle W, \{I_i\}_{i \in \mathcal{A}}, [\![\ ]\!] \rangle$$

where $W$ is a non-empty set of elements called "worlds",
$[\![\ ]\!] : Atm \to \mathcal{P}(W)$ is the evaluation for atomic formulas,
for each $i \in \mathcal{A}$, $I_i : W \to \mathcal{P}(\mathcal{P}(W))$ is the neighbourhood function, satisfying the following properties:

1. **Non-emptiness**: $\forall \phi \in I_i(x), \phi \neq \emptyset$
2. **Nesting**: $\forall \phi, \beta \in I_i(x), \phi \subseteq \beta$ or $\beta \subseteq \phi$
3. **Total reflexivity**: $\exists \phi \in I_i(x)$ such that $x \in \phi$
4. **Local absoluteness**: If $\phi \in I_i(x)$ and $y \in \phi$ then $I_i(x) = I_i(y)$
5. **Closure under intersection**: If $S \subseteq I_i(x)$ and $S \neq \emptyset$ then $\bigcap S \in S$.

We thus require that every neighbourhood of $x$ is nonempty (1), that the neighbourhoods of $x$ are linearly ordered (2), that there is at least one neighbourhood of $x$ that contains $x$ (3), that the neighbourhoods of $x$ and of any element that belongs to one such neighbourhood concide (4), and that we have a closure under intersection (5), a property that would follow from (2) in the finite case but not in general.[7]

A proof of completeness of the axiomatization of the logic of contitional belief with respect to multi-agent canonical models is given in Sect. 2 of Girlando et al. (2018b) using a canonical model construction.

We can now finally define conditional belief in neighbourhood models as:

**Definition 12.3.7** (*Conditional Belief*)

$$x \Vdash Bel_i(B|A) \textit{ iff } \forall \phi \in I_i(x)(\phi \cap [\![A]\!] = \emptyset) \textit{ or } \exists \beta \in I_i(x)(\beta \cap [\![A]\!] \neq \emptyset \textit{ and } \beta \cap [\![A]\!] \subseteq [\![B]\!])$$

This definition can be made more compact using local forcing relations $\Vdash^{\forall}$ and $\Vdash^{\exists}$, introduced in (Negri, 2017b).

**Definition 12.3.8** (*Local forcing relations*)

$$a \Vdash^{\forall} A \textit{ iff } \forall y \in a.y \Vdash A$$
$$a \Vdash^{\exists} A \textit{ iff } \exists y \in a.y \Vdash A$$

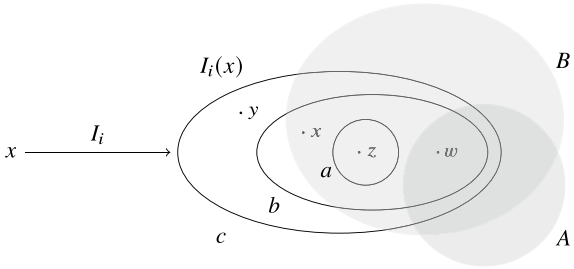
This allows us to then render the final iteration of Definition 12.3.7 as

$$x \Vdash Bel_i(B|A) \textit{ iff } (\forall \phi \in I_i(x).\phi \Vdash^{\forall} \neg A) \textit{ or } \exists \beta \in I_i(x).\beta \Vdash^{\exists} A \textit{ and } \beta \Vdash^{\forall} A \supset B$$

The intuitive meaning of the definition is that an agent has conditional belief in $B$ given $A$ if either $A$ is utterly false (so that we have an inbuilt *ex falso quodlibet*) or there are conceivable sets of epistemic states of the agent where $A$ can be true, and whenever it is true also $B$ is true.

[7] All such properties are required to get the correspondence with plausibility models. However, one can obtain interesting logical systems by giving up some of them. Indeed, this is exactly the reason why neighbourhood semantics is so flexible; see e.g. Negri and Olivetti (2015) and Girlando et al. (2021).

Graphically, the more involved second disjunct can be represented as




Here, there is an element of $I_i(x)$, specifically $b$, such that $b \Vdash^{\exists} A$, since $w \in b$ and $w \Vdash A$, and moreover $b \Vdash^{\forall} A \supset B$, since every element of $b$ (namely $x$, $z$ and $w$) is an element of $(W - [\![A]\!]) \cup B$. The first disjunct could be represented graphically by deleting the circle representing $A$.

#### 12.3.1.1 Relation Between Neighbourhood and Plausibility Models

We briefly establish the adequacy of the approach taken here by recalling the equivalence results between epistemic plausibility models and neighbourhood models, cf. Girlando et al. (2018b).

**Lemma 12.3.9** *If a formula $A$ is valid in the class of epistemic plausibility models then it is valid in the class of multi-agent neighbourhood models.*

***Proof*** Let $\mathcal{M}_P = \langle W, \{\sim_i\}_{i\in\mathcal{A}}, \{\preceq_i\}_{i\in\mathcal{A}}, [\![\ ]\!]\rangle$ be a *plausibility model* ($P$-model).

Let $u \in W$, and define its downward closed set: $\downarrow^{\preceq_i} u = \{v \in W \mid v \preceq_i u\}$.

We define the *neighbourhood model* ($N$-model) $\mathcal{M}_N = \langle W, \{I_i\}_{i\in\mathcal{A}}, [\![\ ]\!]\rangle$, where for $x \in W$ $I_i(x) = \{\downarrow^{\preceq_i} u \mid u \sim_i x\}$.

Trivial for the connectives and straightforward for $Bel_i(B|A)$. □

**Lemma 12.3.10** *If a formula $A$ is valid in the class of multi-agent neighbourhood models then it is valid in the class of epistemic plausibility models.*

***Proof*** Let $\mathcal{M}_N = \langle W, \{I\}_{i\in\mathcal{A}}, [\![\ ]\!]\rangle$ be a multi-agent $N$-model.

We construct a $P$-model $\mathcal{M}_P = \langle W, \{\sim_i\}_{i\in\mathcal{A}}, \{\preceq_i\}_{i\in\mathcal{A}}, [\![\ ]\!]\rangle$, by stipulating:

1. $x \sim_i y$ *iff* $\exists \phi \in I_i(x), y \in \phi$
2. $x \preceq_i y$ *iff* $\forall \phi \in I_i(y)$, if $y \in \phi$ then $x \in \phi$.

Again trivial for the connectives and straightforward for $Bel_i(B|A)$. □

Combined these two lemmas show the following.

**Theorem 12.3.11** *A formula $A$ is a theorem of CDL if and only if it is valid in the class of neighbourhood models.*

## 12.4 Knowledge and Simple Belief

We now move to the discussion of the central concepts of formal epistemology, knowledge and simple belief, which we add to our language (alongside $\top$). We use the local conditional $x \Vdash_i B|A$ to abbreviate $\exists \beta \in I_i(x).\beta \Vdash^{\exists} A \ \wedge \beta \Vdash^{\forall} A \supset B$.

**Definition 12.4.1** (*Language of KSB*)

$$A \ := \ P \mid \bot \mid \top \mid \neg A \mid A \wedge A \mid A \vee A \mid A \supset A \mid Bel_i(A|A) \mid K_i A \mid Bel_i A$$

$K_i A$ and $Bel_i A$ can be defined, as in Girlando et al. (2018b), following Stalnaker (1998), as special cases of conditional belief as follows.

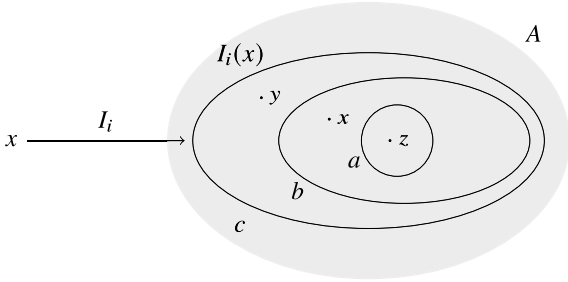
**Definition 12.4.2** (*Knowledge and simple belief in CDL*)
Knowledge: $K_i A \equiv Bel_i(\bot|\neg A)$
Simple belief: $Bel_i A \equiv Bel_i(A|\top)$

That is to say, knowledge is conditional belief that cannot be consistently disputed, while simple belief is conditional belief with a trivial condition. Unpacking these using Definitions 12.3.7 and 12.3.8, we obtain the following truth conditions:

**Definition 12.4.3** (*Knowledge*)

$$\begin{aligned} x \Vdash K_i A \ \textit{iff} \ & \forall \beta \in I_i(x) \ (\beta \subseteq [\![A]\!]) \\ \textit{iff} \ & \forall \beta \in I_i(x) \ (\beta \Vdash^{\forall} A) \end{aligned}$$

Graphically these conditions can be represented as:



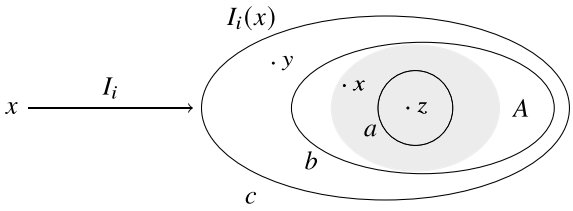

The condition states that *every* neighbourhood in $I_i(x)$ is a subset of $[\![A]\!]$, i.e. every element in the neighbourhood forces $A$.

**Definition 12.4.4** (*Simple belief*)

$$\begin{aligned} x \Vdash Bel_i A \ \textit{iff} \ & \exists \phi \in I_i(x) \ (\phi \subseteq [\![A]\!]) \\ \textit{iff} \ & \exists \phi \in I_i(x) \ (\phi \Vdash^{\forall} A) \end{aligned}$$

The above semantic clause is read off as a special case from the clause for conditional belief, however, it can also be justified independently: in the correspondence between plausibility and neighbourhood models recalled in Sect. 12.3.1.1 above, the neighbourhoods of $x$ for agent $i$ are given by $I_i(x) = \{\downarrow^{\preceq_i} u \mid u \sim_i x\}$, so they are worlds that are the closest, in the plausibility ordering, to worlds epistemically equivalent to $x$. So an agent believes a proposition at world $x$ if the proposition is true in the worlds that are closest to an epistemic state of hers equivalent to $x$.

Graphically these conditions can be represented as:




The condition states that there is *some* neighbourhood of $I_i(x)$ that is a subset of $[\![A]\!]$, i.e. such that every element in the neighbourhood forces $A$.

We proceed by transforming these semantic conditions (specifically the second formulation utilizing the auxiliary local forcing relation) into rules of a sequent calculus.

## 12.4.1 Sequent Calculus G3SBK

To obtain the sequent calculus G3SBK (Table 12.1) we retain the rules of G3CDL, and extend them with rules for simple belief and knowledge, which adhere to Definitions 12.4.3 and 12.4.4. The basic unit of a sequent calculus is a *sequent* $\psi \Rightarrow \Delta$, where $\psi$ and $\Delta$ are finite multisets of (labelled) formulas. All the formulas except $\psi$ and $\Delta$ (which we refer to as the *context*) are called *active* formulas of the rule if they occur only in the upper sequent(s) and *principal* if they occur in the lower sequent of the rule.

The generation of the rules of a labelled sequent calculus follows the methodology established by Negri (2005), of internalizing possible worlds semantics into the syntax of a contraction-free sequent system. When the semantic apparatus is a neighbourhood semantics rather than a standard (relational) possible worlds semantics, the procedure of internalization becomes more articulated. A methodological discussion on the stages needed to find the rules of a well-behaved sequent calculus starting from the meaning explanation of the logical constants in terms of a neighbourhood semantics is given by Negri (2017a).

Here, we just observe that the forcing condition for each connective is an equivalence that gives rise to two rules, a left and a right rule. For conditional belief

**Table 12.1** G3SBK

**Initial sequents** $x : P, \psi \Rightarrow \Delta, x : P$

**Propositional rules:** those of **G3K** (Negri 2005)

**Rules for local forcing**

$$\frac{x \in a, \psi \Rightarrow \Delta, x : A}{\psi \Rightarrow \Delta, a \Vdash^{\forall} A} \; R\Vdash^{\forall} (x \textit{ fresh}) \qquad \frac{x : A, x \in a, a \Vdash^{\forall} A, \psi \Rightarrow \Delta}{x \in a, a \Vdash^{\forall} A, \psi \Rightarrow \Delta} \; L\Vdash^{\forall}$$

$$\frac{x \in a, \psi \Rightarrow \Delta, x : A, a \Vdash^{\exists} A}{x \in a, \psi \Rightarrow \Delta, a \Vdash^{\exists} A} \; R\Vdash^{\exists} \qquad \frac{x \in a, x : A, \psi \Rightarrow \Delta}{a \Vdash^{\exists} A, \psi \Rightarrow \Delta} \; L\Vdash^{\exists} (x \textit{ fresh})$$

**Rules for conditional belief**

$$\frac{a \in I_i(x), a \Vdash^{\exists} A, \psi \Rightarrow \Delta, x \Vdash_i B|A}{\psi \Rightarrow \Delta, x : Bel_i(B|A)} \; RB_i \; (a \textit{ fresh})$$

$$\frac{a \in I_i(x), x : Bel_i(B|A), \psi \Rightarrow \Delta, a \Vdash^{\exists} A \quad x \Vdash_i B|A, a \in I_i(x), x : Bel_i(B|A), \psi \Rightarrow \Delta}{a \in I_i(x), x : Bel_i(B|A), \psi \Rightarrow \Delta} \; LB_i$$

$$\frac{a \in I_i(x), \psi \Rightarrow \Delta, x \Vdash_i B|A, a \Vdash^{\exists} A \quad a \in I_i(x), \psi \Rightarrow \Delta, x \Vdash_i B|A, a \Vdash^{\forall} A \supset B}{a \in I_i(x), \psi \Rightarrow \Delta, x \Vdash_i B|A} \; RC$$

$$\frac{a \in I_i(x), a \Vdash^{\exists} A, a \Vdash^{\forall} A \supset B, \psi \Rightarrow \Delta}{x \Vdash_i B|A, \psi \Rightarrow \Delta} \; LC (a \textit{ fresh})$$

**Rules for inclusion**

$$\frac{a \subseteq a, \psi \Rightarrow \Delta}{\psi \Rightarrow \Delta} \; Ref \qquad \frac{c \subseteq a, c \subseteq b, b \subseteq a, \psi \Rightarrow \Delta}{c \subseteq b, b \subseteq a, \psi \Rightarrow \Delta} \; Tr$$

$$\frac{x \in a, a \subseteq b, x \in b, \psi \Rightarrow \Delta}{x \in a, a \subseteq b, \psi \Rightarrow \Delta} \; L\subseteq$$

**Rules for semantic conditions**

$$\frac{a \subseteq b, a \in I_i(x), b \in I_i(x), \psi \Rightarrow \Delta \quad b \subseteq a, a \in I_i(x), b \in I_i(x), \psi \Rightarrow \Delta}{a \in I_i(x), b \in I_i(x), \psi \Rightarrow \Delta} \; S$$

$$\frac{y \in a, a \in I_i(x), \psi \Rightarrow \Delta}{a \in I_i(x), \psi \Rightarrow \Delta} \; N \; (y \textit{ fresh}) \qquad \frac{x \in a, a \in I_i(x), \psi \Rightarrow \Delta}{\psi \Rightarrow \Delta} \; T \; (a \textit{ fresh})$$

$$\frac{a \in I_i(x), y \in a, b \in I_i(x), b \in I_i(y), \psi \Rightarrow \Delta}{a \in I_i(x), y \in a, b \in I_i(x), \psi \Rightarrow \Delta} \; A_1$$

$$\frac{a \in I_i(x), y \in a, b \in I_i(x), b \in I_i(y), \psi \Rightarrow \Delta}{a \in I_i(x), y \in a, b \in I_i(y), \psi \Rightarrow \Delta} \; A_2$$

$$\frac{a \in I_i(x), y \in a, a \in I_i(y), \psi \Rightarrow \Delta}{a \in I_i(x), y \in a, \psi \Rightarrow \Delta} \; A_1^*$$

**Rules for knowledge and belief**

$$\frac{a \in I_i(x), \psi \Rightarrow \Delta, a \Vdash^{\forall} A}{\psi \Rightarrow \Delta, x : K_i A} \; RK_i \; (a \textit{ fresh}) \qquad \frac{a \in I_i(x), x : K_i A, a \Vdash^{\forall} A, \psi \Rightarrow \Delta}{a \in I_i(x), x : K_i A, \psi \Rightarrow \Delta} \; LK_i$$

$$\frac{a \in I_i(x), \psi \Rightarrow \Delta, x : Bel_i A, a \Vdash^{\forall} A}{a \in I_i(x), \psi \Rightarrow \Delta, x : Bel_i A} \; RSB_i \qquad \frac{a \in I_i(x), a \Vdash^{\forall} A, \psi \Rightarrow \Delta}{x : Bel_i A, \psi \Rightarrow \Delta} \; LSB_i \; (a \textit{ fresh})$$

**Table 12.2** Structural rules

$$\frac{\psi \Rightarrow \Delta}{\mathcal{F}, \psi \Rightarrow \Delta}\ LWeak \qquad \frac{\psi \Rightarrow \Delta}{\psi \Rightarrow \Delta, \mathcal{F}}\ RWeak$$

$$\frac{\mathcal{F}, \mathcal{F}, \psi \Rightarrow \Delta}{\mathcal{F}, \psi \Rightarrow \Delta}\ LCtr \qquad \frac{\psi \Rightarrow \Delta, \mathcal{F}, \mathcal{F}}{\psi \Rightarrow \Delta, \mathcal{F}}\ RCtr$$

$$\frac{\psi \Rightarrow \Delta, \mathcal{F} \quad \mathcal{F}, \psi' \Rightarrow \Delta'}{\psi, \psi' \Rightarrow \Delta, \Delta'}\ Cut$$

the semantic clause is factored through the use of the local conditional $x \Vdash_i B|A$ that therefore appears in the premisses of the rules and has in turn its own rules, $RC$ and $LC$.

Let us look closer at the rules for knowledge and belief: the right rule for knowledge, $RK_i$, has the condition that the neighbourhood variable $a$ should be fresh. So the premiss contains the assumption that for an arbitrary neighbourhood $a \in I_i(x)$, all the worlds in $a$ force $A$. This assumption is next to contexts, i.e. the multisets of formulas $\psi$ and $\Delta$. The rule has the conclusion that $x$ forces $K_i A$.

The left rule for knowledge, in turn, can be read (moving from the antecedent of the conclusion to the antecedent of the premiss), as saying that if $x$ forces $K_i A$ and $a \in I_i(x)$, then $a$ forces universally $A$.

The rules for belief can be seen as dual. The left rule has a freshness condition, since if $x$ forces $Bel_i A$ (antecedent of the conclusion), there is $a \in I_i(x)$ such that $a$ forces universally $A$ (antecedent of the premiss). The right rule is less immediate to justify: the direct semantic condition for $x$ forces $Bel_i A$ would give a rule with two premisses, one with $a \in I_i(x)$, the other with $a \Vdash^{\forall} A$ in their succedents. However, a rule of this form is equivalent to the one written in the table, with only one premiss. The formula $x : Bel_i A$ appears also in the premiss to retain invertibility.

A note on the rule $A_1^*$—it is a contracted version of the rule $A_1$. Since the latter rule has an instance in which two atoms in the conclusion are identical (in the version displayed in the table this would be two instances of $a \in I_i(x)$), in order to allow for contraction, by the closure condition (Negri & von Plato, 2011) the former rule needs to be added.

## 12.5 Structural Properties of G3SBK

The structural properties of the segment of the system without the rules for knowledge and belief have been demonstrated by Girlando et al. (2018b). Here we extend those proofs for the added rules. The structural rules are summarized in Table 12.2.

To start with, we observe that the usual substitution results hold unaltered. The definition of substitution of labels given by Negri (2005) is extended in an obvious way—that need not be detailed here—to all the formulas of our language and to neighbourhood labels. With this definition we have, for example, $(a \Vdash^{\exists} A)(b/a) \equiv b \Vdash^{\exists} A$, and $(x \Vdash_i B|A)(y/x) \equiv y \Vdash_i B|A$.

We recall that the *height* of a derivation is its height as a tree, i.e., the length of its longest branch, and that $\vdash_n$ denotes derivability with derivation height bounded by $n$ in a given system.

The calculus is shown, by a routine induction on the height of derivations, to enjoy the property of height-preserving substitution both of world and neighbourhood labels. Details are found in (Girlando et al., 2018b).

**Proposition 12.5.1** *(i) If* $\vdash_n \psi \Rightarrow \Delta$, *then* $\vdash_n \psi(y/x) \Rightarrow \Delta(y/x)$; *(ii) If* $\vdash_n \psi \Rightarrow \Delta$, *then* $\vdash_n \psi(b/a) \Rightarrow \Delta(b/a)$.

For the other properties we need a finer induction that takes into account also the weight of formulas, that we extend from the one in (Girlando et al., 2018b).

**Definition 12.5.2** The *weight* of the labelled formula $\mathcal{F}$ is the pair

$$(\mathtt{w}(p(\mathcal{F})), \mathtt{w}(l(\mathcal{F})))$$

where $l(\mathcal{F})$ is the label of $\mathcal{F}$, with

$\mathtt{w}(x) = 0,$
$\mathtt{w}(a) = 1,$

and $p(\mathcal{F})$ is the part of $\mathcal{F}$ without the label and the forcing relation, with

$\mathtt{w}(P) = \mathtt{w}(\bot) = \mathtt{w}(\top) = 1,$
$\mathtt{w}(A \circ B) = \mathtt{w}(A) + \mathtt{w}(B) + 1, \circ \in \{\vee, \,\&\, , \supset\}$
$\mathtt{w}(\neg A) = \mathtt{w}(A) + 2,$
$\mathtt{w}(B|A) = \mathtt{w}(A) + \mathtt{w}(B) + 2,$
$\mathtt{w}(Bel_i(B|A)) = \mathtt{w}(B|A) + 1,$
$\mathtt{w}(Bel_i A) = \mathtt{w}(A) + 4,$
$\mathtt{w}(K_i A) = \mathtt{w}(A) + 6.$

Weights of labelled formulas are ordered lexicographically.

Observe that a formula of the form $a \Vdash^{\forall} A$ is a labelled formula, with $a$ its label and $A$ the part without label. This definition of weights unfolds those for $Bel_i$ and $K_i$ in accordance with Definition 12.4.2.

**Lemma 12.5.3** *For any labelled formula* $\mathcal{F}$, *the sequent* $\mathcal{F}, \psi \Rightarrow \Delta, \mathcal{F}$ *is derivable.*

***Proof*** By induction on the weight of the formula. We need to check the new cases, for $x : Bel_i A, \psi \Rightarrow \Delta, x : Bel_i A$ and $x : K_i A, \psi \Rightarrow \Delta, x : K_i A$

$$\dfrac{\dfrac{a \in I_i(x), a \Vdash^{\forall} A, \psi \Rightarrow \Delta, x : Bel_i A, a \Vdash^{\forall} A}{a \in I_i(x), a \Vdash^{\forall} A, \psi \Rightarrow \Delta, x : Bel_i A}\, RSB_i}{x : Bel_i A, \psi \Rightarrow \Delta, x : Bel_i A}\, LSB_i$$

The top sequent is derivable by the inductive hypothesis, since we have a smaller weight, $\mathtt{w}(a \Vdash^{\forall} A) < \mathtt{w}(x : Bel_i A)$ (recall that the weights are ordered lexicographically).

$$\dfrac{\dfrac{a \in I_i(x), x : K_i A, a \Vdash^{\forall} A, \psi \Rightarrow \Delta, a \Vdash^{\forall} A}{a \in I_i(x), x : K_i A, \psi \Rightarrow \Delta, a \Vdash^{\forall} A}\, LK_i}{x : K_i A, \psi \Rightarrow \Delta, x : K_i A}\, RK_i$$

The top sequent is again derivable by the inductive hypothesis, since we have $\mathtt{w}(a \Vdash^{\forall} A) < \mathtt{w}(x : K_i A)$. □

**Lemma 12.5.4** *Weakening is height-preserving admissible.*

***Proof*** Routine by induction on the height of the derivation. □

We move on to demonstrate invertibility of the rules. The only rules we need to address are the rules added to G3CDL.

**Lemma 12.5.5** *The following hold:*

1. *If* $\vdash_n a \in I_i(x), \psi \Rightarrow \Delta$ *then* $\vdash_n y \in a, a \in I_i(x), \psi \Rightarrow \Delta$
2. *If* $\vdash_n a \in I_i(x), y \in a, b \in I_i(y), \psi \Rightarrow \Delta$ *then* $\vdash_n a \in I_i(x), y \in a, b \in I_i(x), b \in I_i(y), \psi \Rightarrow \Delta$
3. *If* $\vdash_n \psi \Rightarrow \Delta, x : K_i A$ *then* $\vdash_n a \in I_i(x), \psi \Rightarrow \Delta, a \Vdash^{\forall} A$
4. *If* $\vdash_n a \in I_i(x), x : K_i A, \psi \Rightarrow \Delta$ *then* $\vdash_n a \in I_i(x), x : K_i A, a \Vdash^{\forall} A, \psi \Rightarrow \Delta$
5. *If* $\vdash_n a \in I_i(x), \psi \Rightarrow \Delta, x : Bel_i A$ *then* $\vdash_n a \in I_i(x), \psi \Rightarrow \Delta, x : Bel_i A, a \Vdash^{\forall} A$
6. *If* $\vdash_n x : Bel_i A, \psi \Rightarrow \Delta$ *then* $\vdash_n a \in I_i(x), a \Vdash^{\forall} A, \psi \Rightarrow \Delta$.

***Proof*** (1), (2), (4) and (5) are instances of Lemma 12.5.4. The remaining proofs are by induction on the height of the derivation.

(3) Base case. Suppose $\psi \Rightarrow \Delta, x : K_i A$ is an initial sequent or a conclusion of L⊥. In the former case, since $x : K_i A$ is neither atomic nor a relational atom, so is $\vdash_n a \in I_i(x), \psi \Rightarrow \Delta, a \Vdash^{\forall} A$, and in the latter we have that it is a conclusion of L⊥.

Inductive case. If the last step of the derivation is some rule $R$ other than $\mathrm{RK}_i$, or it is $\mathrm{RK}_i$ with some other principal formula, then it is of the form

$$\dfrac{\psi' \Rightarrow \Delta', x : K_i A \quad [\psi'' \Rightarrow \Delta'', x : K_i A]}{\psi \Rightarrow \Delta, x : K_i A}\, R$$

We apply the inductive hypothesis to the premiss(es), first applying substitution if necessary, and then reapplying the rule $R$ to obtain the derivation

$$\dfrac{a \in I_i(x), \psi' \Rightarrow \Delta', a \Vdash^{\forall} A \quad [a \in I_i(x), \psi'' \Rightarrow \Delta'', a \Vdash^{\forall} A]}{a \in I_i(x), \psi \Rightarrow \Delta, a \Vdash^{\forall} A}\, R$$

If the last rule applied is $\mathrm{RK}_i$, then its premiss is already the requisite sequent.

(6) The proof exactly mirrors, *mutatis mutandis*, the proof for (3). □

**Lemma 12.5.6** *The rules of left and right contraction are height-preserving admissible.*

***Proof*** By simultaneous induction on the height of the derivation.

Base case. If the sequent under consideration is an initial sequent or a conclusion of the rule L$\bot$ then so is the contracted sequent.

Inductive case. If the contracted formula is not principal in the last rule, we apply the inductive hypothesis to its premiss(es) and then reapply the rule to obtain the requisite sequent. Otherwise we only need to check the new rules. If the last step of the derivation is the rule $RK_i$, then it is of the form

$$\frac{a \in I_i(x), \psi \Rightarrow \Delta, x : K_i A, a \Vdash^{\forall} A}{\psi \Rightarrow \Delta, x : K_i A, x : K_i A} \; RK_i$$

Using Lemma 12.5.5, we obtain the sequent $a \in I_i(x), a \in I_i(x), \psi \Rightarrow \Delta, a \Vdash^{\forall} A, a \Vdash^{\forall} A$. Since it is of lower height, we apply the inductive hypothesis and then the rule $RK_i$ to obtain the sequent $\psi \Rightarrow \Delta, x : K_i A$. Similar for $LSB_i$. If the last step is one of the rules $N$, $A_2$, $LK_i$ or $RSB_i$, we apply the inductive hypothesis to the premiss and then reapply the rule. □

**Theorem 12.5.7** *Cut is admissible.*

***Proof*** By primary induction on the weight of the formula and secondary induction on the sum of the heights of the premisses of cut. The only cases we need to inspect are those where the cut formula is principal in both premisses and of the form $x : K_i A$ or $x : Bel_i A$ because the other cases are either treated by Girlando et al. (2018b) or are inessential modifications of cases treated there, in which the cut is reduced via a permutation upwards. In the former case we have a derivation

$$\dfrac{\dfrac{b \in I_i(x), \psi \Rightarrow \Delta, b \Vdash^{\forall} A}{\psi \Rightarrow \Delta, x : K_i A} \; RK_i \qquad \dfrac{a \in I_i(x), x : K_i A, a \Vdash^{\forall} A, \psi' \Rightarrow \Delta'}{a \in I_i(x), x : K_i A, \psi' \Rightarrow \Delta'} \; LK_i}{a \in I_i(x), \psi', \psi \Rightarrow \Delta, \Delta'} \; Cut$$

This is transformed into:

$$\dfrac{\dfrac{\dfrac{b \in I_i(x), \psi \Rightarrow \Delta, b \Vdash^{\forall} A}{a \in I_i(x), \psi \Rightarrow \Delta, a \Vdash^{\forall} A} \; Lm\ 12.5.1 \qquad \dfrac{\dfrac{b \in I_i(x), \psi \Rightarrow \Delta, b \Vdash^{\forall} A}{\psi \Rightarrow \Delta, x : K_i A} \; RK_i \qquad a \in I_i(x), x : K_i A, a \Vdash^{\forall} A, \psi' \Rightarrow \Delta'}{a \in I_i(x), a \Vdash^{\forall} A, \psi', \psi \Rightarrow \Delta, \Delta'} \; Cut_1}{a \in I_i(x), a \in I_i(x), \psi', \psi, \psi \Rightarrow \Delta, \Delta, \Delta'} \; Cut_2}{a \in I_i(x), \psi', \psi \Rightarrow \Delta, \Delta'} \; Lm\ 12.5.6$$

The application of the Cut rule labelled $Cut_1$ is of lower height, and that labelled $Cut_2$ is of lower weight (recall the lexicographical ordering).

If the cut formula is of the form $x : Bel_i A$ we have a derivation

$$\dfrac{\dfrac{a \in I_i(x), \psi \Rightarrow \Delta, x : Bel_i A, a \Vdash^{\forall} A}{a \in I_i(x), \psi \Rightarrow \Delta, x : Bel_i A}\, RSB_i \qquad \dfrac{b \in I_i(x), b \Vdash^{\forall} A, \psi' \Rightarrow \Delta'}{x : Bel_i A, \psi' \Rightarrow \Delta'}\, LSB_i}{a \in I_i(x), \psi', \psi \Rightarrow \Delta, \Delta'}\, Cut$$

This is transformed into:

$$\dfrac{\dfrac{\dfrac{a \in I_i(x), \psi \Rightarrow \Delta, x : Bel_i A, a \Vdash^{\forall} A \qquad \dfrac{b \in I_i(x), b \Vdash^{\forall} A, \psi' \Rightarrow \Delta'}{x : Bel_i A, \psi' \Rightarrow \Delta'}\, LSB_i}{a \in I_i(x), \psi', \psi \Rightarrow \Delta, \Delta', a \Vdash^{\forall} A}\, Cut_1 \qquad \dfrac{b \in I_i(x), b \Vdash^{\forall} A, \psi' \Rightarrow \Delta'}{a \in I_i(x), a \Vdash^{\forall} A, \psi' \Rightarrow \Delta'}\, Prop\ 12.5.1}{a \in I_i(x), a \in I_i(x), \psi', \psi' \Rightarrow \Delta, \Delta', \Delta'}\, Cut_2}{a \in I_i(x), \psi', \psi \Rightarrow \Delta, \Delta'}\, Lm\ 12.5.6$$

Again the application of the Cut rule labelled $Cut_1$ is of lower height, and that labelled $Cut_2$ is of lower weight. □

## 12.6 Relationship to G3CDL

Under Definition 12.4.2, the rules for knowledge and simple belief are admissible in CDL (Girlando et al., 2018b). This has been shown for the case of the rule $\mathrm{LK}_i$. We demonstrate this result for the remaining cases, thereby also illustrating the need for the semantic conditions for neighbourhood frames, without which the calculus wouldn't be complete.

We begin by proving several auxiliary results useful for abbreviating (relying on admissibility of cut) more involved derivations. Observe that in the proofs of Lemmas 12.6.1 and 12.6.2 below we use that top sequents are derivable by Lemma 12.5.3.

**Lemma 12.6.1** *The local forcing relations* $\Vdash^{\forall}$ *and* $\Vdash^{\exists}$ *obey the De Morgan laws, i.e. the following sequents are derivable:*

$(i)\ a \Vdash^{\exists} A, a \Vdash^{\forall} \neg A \Rightarrow$ $\qquad$ $(ii)\ \Rightarrow a \Vdash^{\exists} A, a \Vdash^{\forall} \neg A$

$(iii)\ a \Vdash^{\forall} A, a \Vdash^{\exists} \neg A \Rightarrow$ $\qquad$ $(iv)\ \Rightarrow a \Vdash^{\forall} A, a \Vdash^{\exists} \neg A$

***Proof***

(i)

$$\dfrac{\dfrac{\dfrac{x \in a, x : A, a \Vdash^{\forall} \neg A \Rightarrow x : A}{x : \neg A, x \in a, x : A, a \Vdash^{\forall} \neg A \Rightarrow}\, L\neg}{x \in a, x : A, a \Vdash^{\forall} \neg A \Rightarrow}\, L\Vdash^{\forall}}{a \Vdash^{\exists} A, a \Vdash^{\forall} \neg A \Rightarrow}\, L\Vdash^{\exists}$$

(ii)

$$\dfrac{\dfrac{\dfrac{x:A, x\in a \Rightarrow a\Vdash^{\exists} A, x:A}{x:A, x\in a \Rightarrow a\Vdash^{\exists} A}\, R\Vdash^{\exists}}{x\in a \Rightarrow a\Vdash^{\exists} A, x:\neg A}\, R\neg}{\Rightarrow a\Vdash^{\exists} A, a\Vdash^{\forall}\neg A}\, R\Vdash^{\forall}$$

(iii) and (iv) are very similar to (i) and (ii), respectively. □

**Lemma 12.6.2** $a\Vdash^{\forall}\top\supset A$ if and only if a$\Vdash^{\forall}$A.

***Proof***

$$\dfrac{\dfrac{\dfrac{}{x\in a, a\Vdash^{\forall}\top\supset A\Rightarrow x:A, x:\top}\, R\top \quad x:A, x\in a, a\Vdash^{\forall}\top\supset A\Rightarrow x:A}{x:\top\supset A, x\in a, a\Vdash^{\forall}\top\supset A\Rightarrow x:A}\, L\supset}{x\in a, a\Vdash^{\forall}\top\supset A\Rightarrow x:A}\, L\Vdash^{\forall}}{a\Vdash^{\forall}\top\supset A\Rightarrow a\Vdash^{\forall} A}\, R\Vdash^{\forall}$$

$$\dfrac{\dfrac{\dfrac{x:\top, x\in a, a\Vdash^{\forall} A, x:A\Rightarrow x:A}{x:\top, x\in a, a\Vdash^{\forall} A\Rightarrow x:A}\, L\Vdash^{\forall}}{x\in a, a\Vdash^{\forall} A\Rightarrow x:\top\supset A}\, R\supset}{a\Vdash^{\forall} A\Rightarrow a\Vdash^{\forall}\top\supset A}\, R\Vdash^{\forall}$$

□

**Lemma 12.6.3** *The translation of the rule $RK_i$ according to Definition 12.4.2 is admissible in G3CDL.*

***Proof***

$$\dfrac{\dfrac{a\in I_i(x), a\Vdash^{\exists}\neg A, \psi\Rightarrow\Delta, x\Vdash_i \bot|\neg A, a\Vdash^{\exists}\neg A \quad \dfrac{\dfrac{a\in I_i(x), \psi\Rightarrow\Delta, a\Vdash^{\forall} A}{a\in I_i(x), \psi\Rightarrow\Delta, x\Vdash_i \bot|\neg A, a\Vdash^{\forall}\neg A\supset\bot, a\Vdash^{\forall} A}\, W}{a\in I_i(x), a\Vdash^{\exists}\neg A, \psi\Rightarrow\Delta, x\Vdash_i \bot|\neg A, a\Vdash^{\forall}\neg A\supset\bot}\, Lm\ 12.6.1}{a\in I_i(x), a\Vdash^{\exists}\neg A, \psi\Rightarrow\Delta, x\Vdash_i \bot|\neg A}\, RC}{\psi\Rightarrow\Delta, x:Bel_i(\bot|\neg A)}\, RB_i$$

□

**Lemma 12.6.4** *The translation of the rule $LSB_i$ according to Definition 12.4.2 is admissible in G3CDL.*

***Proof***

(1)

$$\dfrac{\dfrac{\dfrac{a\in I_i(x), a\Vdash^{\forall} A, \psi\Rightarrow\Delta}{a\in I_i(x), a\Vdash^{\exists}\top, a\Vdash^{\forall} A, x\in b, b\in I_i(x), x:Bel_i(A|\top), \psi\Rightarrow\Delta}\, W}{a\in I_i(x), a\Vdash^{\exists}\top, a\Vdash^{\forall}\top\supset A, x\in b, b\in I_i(x), x:Bel_i(A|\top), \psi\Rightarrow\Delta}\, Lm\ 12.6.2}{x\Vdash_i A|\top, x\in b, b\in I_i(x), x:Bel_i(A|\top), \psi\Rightarrow\Delta}\, LC$$

$$\dfrac{\dfrac{\dfrac{}{x \in b, b \in I_i(x), x : Bel_i(A|\top), \psi \Rightarrow \Delta, b \Vdash^{\exists} \top, x : \top}\, R\top}{x \in b, b \in I_i(x), x : Bel_i(A|\top), \psi \Rightarrow \Delta, b \Vdash^{\exists} \top}\, R\Vdash^{\exists} \quad (1)}{x \in b, b \in I_i(x), x : Bel_i(A|\top), \psi \Rightarrow \Delta}\, LB_i$$
$$x : Bel_i(A|\top), \psi \Rightarrow \Delta \; T$$

□

**Lemma 12.6.5** *The translation of the rule $RSB_i$ according to Definition 12.4.2 is admissible in G3CDL.*

***Proof***

(1)

$$\dfrac{\dfrac{a \in I_i(x), \psi \Rightarrow \Delta, x : Bel_i(A|\top), a \Vdash^{\forall} A}{b \in I_i(x), a \in I_i(x), \psi \Rightarrow \Delta, x : Bel_i(A|\top), x \Vdash_i A|\top, a \Vdash^{\forall} A}\, W}{b \in I_i(x), a \in I_i(x), \psi \Rightarrow \Delta, x : Bel_i(A|\top), x \Vdash_i A|\top, a \Vdash^{\forall} \top \supset A}\, Lm\ 12.6.2$$

$$\dfrac{\dfrac{\dfrac{\dfrac{\dfrac{}{y \in a, b \in I_i(x), a \in I_i(x), \psi \Rightarrow \Delta, x : Bel_i(A|\top), x \Vdash_i A|\top, a \Vdash^{\exists} \top, y : \top}\, R\top}{y \in a, b \in I_i(x), a \in I_i(x), \psi \Rightarrow \Delta, x : Bel_i(A|\top), x \Vdash_i A|\top, a \Vdash^{\exists} \top}\, R\Vdash^{\exists}}{b \in I_i(x), a \in I_i(x), \psi \Rightarrow \Delta, x : Bel_i(A|\top), x \Vdash_i A|\top, a \Vdash^{\exists} \top}\, N \quad (1)}{b \in I_i(x), a \in I_i(x), \psi \Rightarrow \Delta, x : Bel_i(A|\top), x \Vdash_i A|\top}\, RC}{a \in I_i(x), \psi \Rightarrow \Delta, x : Bel_i(A|\top), x : Bel_i(A|\top)}\, RB_i$$
$$a \in I_i(x), \psi \Rightarrow \Delta, x : Bel_i(A|\top) \; C$$

□

### 12.6.1 Further Properties of $K_i$: Factivity and Introspection

Since Hintikka's influential monograph (Hintikka, 1962), the classical account of epistemic logic validates three standard properties. The first, called *factivity*, states that whatever is known by an agent must be true, axiomatically expressed by $K_i A \supset A$; the second, *positive introspection*, is the property stating that a knowing agent is aware of her knowledge, i.e. $K_i A \supset K_i K_i A$; finally the third (and most controversial) property, *negative introspection*, states that a knowing agent is aware of her ignorance, formally $\neg K_i A \supset K_i(\neg K_i A)$.

The three properties are those characterizing an S5 modal operator and correspond, in the Kripke semantics, to the accessibility relation being an equivalence relation; see e.g. Van Ditmarsch et al. (2008, Chap. 2) for a compact recent presentation of epistemic logic. We can prove that they also hold for the knowledge operator defined in the present setting, something stated without proof by Girlando et al. (2018b).

Moreover, one can see from the proofs how these properties rely on the semantic conditions, which will be informative in the last section of this paper.

**Theorem 12.6.6** *$K_i$ is an S5 operator, i.e. it satisfies the following axioms:*
(i) $K_i A \supset A$
(ii) $K_i A \supset K_i K_i A$
(iii) $\neg K_i A \supset K_i \neg K_i A$

***Proof***

(i)

$$\dfrac{\dfrac{\dfrac{x \in a, a \in I_i(x), x : K_i A, a \Vdash^{\forall} A, x : A \Rightarrow x : A}{x \in a, a \in I_i(x), x : K_i A, a \Vdash^{\forall} A \Rightarrow x : A}\, L\Vdash^{\forall}}{x \in a, a \in I_i(x), x : K_i A \Rightarrow x : A}\, LK_i}{x : K_i A \Rightarrow x : A}\, T$$

(ii)

$$\dfrac{\dfrac{\dfrac{\dfrac{\dfrac{\dfrac{\dfrac{z : A, b \Vdash^{\forall} A, b \in I_i(x), z \in b, b \in I_i(y), y \in a, a \in I_i(x), x : K_i A \Rightarrow z : A}{b \Vdash^{\forall} A, b \in I_i(x), z \in b, b \in I_i(y), y \in a, a \in I_i(x), x : K_i A \Rightarrow z : A}\, L\Vdash^{\forall}}{b \in I_i(x), z \in b, b \in I_i(y), y \in a, a \in I_i(x), x : K_i A \Rightarrow z : A}\, LK_i}{z \in b, b \in I_i(y), y \in a, a \in I_i(x), x : K_i A \Rightarrow z : A}\, A_2}{b \in I_i(y), y \in a, a \in I_i(x), x : K_i A \Rightarrow b \Vdash^{\forall} A}\, R\Vdash^{\forall}}{y \in a, a \in I_i(x), x : K_i A \Rightarrow y : K_i A}\, RK_i}{a \in I_i(x), x : K_i A \Rightarrow a \Vdash^{\forall} K_i A}\, R\Vdash^{\forall}}{x : K_i A \Rightarrow x : K_i K_i A}\, RK_i$$

(iii)

$$\dfrac{\dfrac{\dfrac{\dfrac{\dfrac{\dfrac{\dfrac{\dfrac{\dfrac{\dfrac{a \in I_i(z), z : K_i A, a \Vdash^{\forall} A, y : A, z \in b, y \in a, b \in I_i(x), a \in I_i(x) \Rightarrow y : A}{a \in I_i(z), z : K_i A, a \Vdash^{\forall} A, z \in b, y \in a, b \in I_i(x), a \in I_i(x) \Rightarrow y : A}\, L\Vdash^{\forall}}{a \in I_i(z), z : K_i A, z \in b, y \in a, b \in I_i(x), a \in I_i(x) \Rightarrow y : A}\, LK_i}{z : K_i A, z \in b, y \in a, b \in I_i(x), a \in I_i(x) \Rightarrow y : A}\, A_1}{z \in b, y \in a, b \in I_i(x), a \in I_i(x) \Rightarrow z : \neg K_i A, y : A}\, R\neg}{y \in a, b \in I_i(x), a \in I_i(x) \Rightarrow b \Vdash^{\forall} \neg K_i A, y : A}\, R\Vdash^{\forall}}{b \in I_i(x), a \in I_i(x) \Rightarrow b \Vdash^{\forall} \neg K_i A, a \Vdash^{\forall} A}\, R\Vdash^{\forall}}{a \in I_i(x) \Rightarrow x : K_i \neg K_i A, a \Vdash^{\forall} A}\, RK_i}{\Rightarrow x : K_i \neg K_i A, x : K_i A}\, RK_i}{x : \neg K_i A \Rightarrow x : K_i \neg K_i A}\, L\neg$$

□

### 12.6.2 The Gettier Problem

We now illustrate what goes wrong in the cases like the ones from Gettier's seminal (Gettier, 1963), this time taking a look at case 2. In it, Smith knows that

(C) Jones has owned a car for as long as Smith can remember, and always a Ford, and that Jones has only recently offered him a ride in a Ford.

So (C) is strong evidence for the claim

(A) Jones owns a Ford.

Picking a place at random, he combines it with a claim

(B) Brown is in Barcelona

to come to believe that, based on (C),

(A∨B) Jones owns a Ford or Brown is in Barcelona.

Given the strength of evidence (C) and propositional logic, Smith is highly justified in believing this. Moreover, this actually happens to be true, although unbeknownst to Smith, Jones is at the moment merely renting a Ford, while, through sheer coincidence, Brown does at the moment find herself in Barcelona. So, it is the second disjunct that accounts for the truth of this conclusion.

Gettier wraps up the brief report by pointing out that, even though the belief is true and justified, it does not count as knowledge.

We formalize the situation as follows: Smith knows $C$, the evidence set that warrants concluding $A$, namely $K_sC$. Moreover, given $C$ there is good reason to hold that $A$, namely $x \Vdash_i A|C$ (there is a situation epistemically indistinguishable from Smith's in which $C$ non-trivially warrants $A$). From this he arrives at a (i) justified (in the sense of there being good reason to hold it as before) (ii) true (iii) belief that $A \vee B$. We show these three points in turn ($L\supset$ at the top omitted).

(i)

$$\dfrac{\dfrac{\vdots}{\dfrac{\dfrac{y \in a, a \in I_s(x), ..., a \Vdash^{\forall} C \supset A, y : C \supset A, y : C, ... \Rightarrow ..., y : A \vee B}{y \in a, a \in I_s(x), ..., a \Vdash^{\forall} C \supset A, y : C \supset A, ... \Rightarrow ..., y : C \supset (A \vee B)}\, R\supset}{y \in a, a \in I_s(x), ..., a \Vdash^{\forall} C \supset A, ... \Rightarrow ..., y : C \supset (A \vee B)}\, L\Vdash^{\forall}}}{\dfrac{\vdots}{..., a \Vdash^{\exists} C, ... \Rightarrow ..., a \Vdash^{\exists} C}\quad \dfrac{}{a \in I_s(x), x : K_sC, a \Vdash^{\exists} C, a \Vdash^{\forall} C \supset A, x : \neg A, x : B \Rightarrow ..., a \Vdash^{\forall} C \supset (A \vee B)}\, R\Vdash^{\forall}}\, RC}{\dfrac{a \in I_s(x), x : K_sC, a \Vdash^{\exists} C, a \Vdash^{\forall} C \supset A, x : \neg A, x : B \Rightarrow x \Vdash_s A \vee B \mid C}{x : K_sC, x \Vdash_s A|C, x : \neg A, x : B \Rightarrow x \Vdash_s A \vee B \mid C}\, LC}$$

(ii) is trivial due to $B$:

$$\dfrac{x : K_sC, x \Vdash_s A|C, x : \neg A, x : B \Rightarrow x : A, x : B}{x : K_sC, x \Vdash_s A|C, x : \neg A, x : B \Rightarrow x : A \vee B}\, R\vee$$

(iii)

$$\dfrac{\dfrac{\dfrac{\dfrac{\dfrac{\dfrac{\dfrac{\vdots}{y \in a, a \in I_s(x), ..., y\colon C, a \Vdash^{\exists} C, ..., y\colon C \supset A, x\colon \neg A, x\colon B \Rightarrow ..., y\colon A \vee B}\, L\Vdash^{\forall}}{y \in a, a \in I_s(x), ..., y\colon C, a \Vdash^{\exists} C, a \Vdash^{\forall} C \supset A, x\colon \neg A, x\colon B \Rightarrow ..., y\colon A \vee B}\, L\Vdash^{\forall}}{y \in a, a \in I_s(x), ..., a \Vdash^{\forall} C, a \Vdash^{\exists} C, a \Vdash^{\forall} C \supset A, x\colon \neg A, x\colon B \Rightarrow ..., y\colon A \vee B}\, R\Vdash^{\forall}}{a \in I_s(x), ..., a \Vdash^{\forall} C, a \Vdash^{\exists} C, a \Vdash^{\forall} C \supset A, x\colon \neg A, x\colon B \Rightarrow ..., a \Vdash^{\forall} A \vee B}\, RBel_s}{a \in I_s(x), ..., a \Vdash^{\forall} C, a \Vdash^{\exists} C, a \Vdash^{\forall} C \supset A, x\colon \neg A, x\colon B \Rightarrow x\colon Bel_s(A \vee B)}\, LK_s}{a \in I_s(x), x\colon K_s C, a \Vdash^{\exists} C, a \Vdash^{\forall} C \supset A, x\colon \neg A, x\colon B \Rightarrow x\colon Bel_s(A \vee B)}\, LC}{x\colon K_s C, x \Vdash_s A | C, x\colon \neg A, x\colon B \Rightarrow x\colon Bel_s(A \vee B)}$$

We now try to construct a derivation of the stronger conclusion of knowledge, $K_s(A \vee B)$, and let the failure instruct us in what went wrong. We break the derivation into several subderivations for legibility.

(1)

$$\dfrac{y\colon C, z\colon A, z \in b, z \in a, z\colon C, a \subseteq b, y \in b, b \in I_s(x), x\colon K_s C, b \Vdash^{\forall} C, a \in I_s(x), a \Vdash^{\forall} C \supset A, ... \Rightarrow y\colon A, y\colon B}{z\colon A, z \in b, z \in a, z\colon C, a \subseteq b, y \in b, b \in I_s(x), x\colon K_s C, b \Vdash^{\forall} C, a \in I_s(x), a \Vdash^{\forall} C \supset A, ... \Rightarrow y\colon A, y\colon B}\, L\Vdash^{\forall}$$

(2)

$$\dfrac{\dfrac{\dfrac{\dfrac{\dfrac{..., z\colon C, ... \Rightarrow y\colon A, y\colon B, z\colon C \qquad (1)}{z\colon C \supset A, z \in b, z \in a, z\colon C, a \subseteq b, y \in b, b \in I_s(x), x\colon K_s C, b \Vdash^{\forall} C, a \in I_s(x), a \Vdash^{\forall} C \supset A, ... \Rightarrow y\colon A, y\colon B}\, L\supset}{z \in b, z \in a, z\colon C, a \subseteq b, y \in b, b \in I_s(x), x\colon K_s C, b \Vdash^{\forall} C, a \in I_s(x), a \Vdash^{\forall} C \supset A, ... \Rightarrow y\colon A, y\colon B}\, L\Vdash^{\forall}}{z \in b, z \in a, z\colon C, a \subseteq b, y \in b, b \in I_s(x), x\colon K_s C, a \in I_s(x), a \Vdash^{\forall} C \supset A, ... \Rightarrow y\colon A, y\colon B}\, LK_s}{z \in a, z\colon C, a \subseteq b, y \in b, b \in I_s(x), x\colon K_s C, a \in I_s(x), a \Vdash^{\forall} C \supset A, ... \Rightarrow y\colon A, y\colon B}\, L\subseteq}{a \subseteq b, y \in b, b \in I_s(x), x\colon K_s C, a \in I_s(x), a \Vdash^{\exists} C, a \Vdash^{\forall} C \supset A, ... \Rightarrow y\colon A, y\colon B}\, L\Vdash^{\exists}$$

(3)

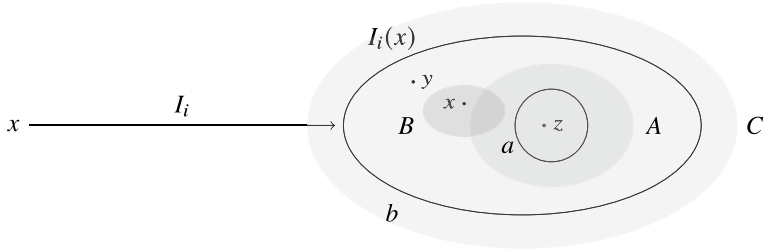
$$\dfrac{\dfrac{\dfrac{\dfrac{\dfrac{\vdots}{y\colon C \supset A, y\colon C, y \in a, b \subseteq a, y \in b, b \in I_s(x), x\colon K_s C, a \Vdash^{\forall} C, a \in I_s(x), a \Vdash^{\exists} C, a \Vdash^{\forall} C \supset A, ... \Rightarrow y\colon A, y\colon B}\, L\Vdash^{\forall}}{y\colon C, y \in a, b \subseteq a, y \in b, b \in I_s(x), x\colon K_s C, a \Vdash^{\forall} C, a \in I_s(x), a \Vdash^{\exists} C, a \Vdash^{\forall} C \supset A, ... \Rightarrow y\colon A, y\colon B}\, L\Vdash^{\forall}}{y \in a, b \subseteq a, y \in b, b \in I_s(x), x\colon K_s C, a \Vdash^{\forall} C, a \in I_s(x), a \Vdash^{\exists} C, a \Vdash^{\forall} C \supset A, ... \Rightarrow y\colon A, y\colon B}\, LK_s}{y \in a, b \subseteq a, y \in b, b \in I_s(x), x\colon K_s C, a \in I_s(x), a \Vdash^{\exists} C, a \Vdash^{\forall} C \supset A, ... \Rightarrow y\colon A, y\colon B}\, L\subseteq}{b \subseteq a, y \in b, b \in I_s(x), x\colon K_s C, a \in I_s(x), a \Vdash^{\exists} C, a \Vdash^{\forall} C \supset A, ... \Rightarrow y\colon A, y\colon B}$$

$$\dfrac{\dfrac{\dfrac{\dfrac{\dfrac{(2) \qquad (3)}{y \in b, b \in I_s(x), x\colon K_s C, a \in I_s(x), a \Vdash^{\exists} C, a \Vdash^{\forall} C \supset A, x\colon \neg A, x\colon B \Rightarrow y\colon A, y\colon B}\, S}{y \in b, b \in I_s(x), x\colon K_s C, a \in I_s(x), a \Vdash^{\exists} C, a \Vdash^{\forall} C \supset A, x\colon \neg A, x\colon B \Rightarrow y\colon A \vee B}\, R\vee}{b \in I_s(x), x\colon K_s C, a \in I_s(x), a \Vdash^{\exists} C, a \Vdash^{\forall} C \supset A, x\colon \neg A, x\colon B \Rightarrow b \Vdash^{\forall} A \vee B}\, R\Vdash^{\forall}}{x\colon K_s C, a \in I_s(x), a \Vdash^{\exists} C, a \Vdash^{\forall} C \supset A, x\colon \neg A, x\colon B \Rightarrow x\colon K_s(A \vee B)}\, RK_s}{x\colon K_s C, x \Vdash_i A | C, x\colon \neg A, x\colon B \Rightarrow x\colon K_s(A \vee B)}\, LC$$

By total reflexivity $x$ is in some element of $I_s(x)$, but rather than complicate the picture further by introducing another element, we situate it according to the formulas it forces. We likewise omit the application of some $A$ rules since we are not dealing with embedded modalities.

We now use the branch ending in (1) to construct a countermodel.

| | | |
|---|---|---|
| $I_s(x) = \{a, b\}$ | $x \nVdash A$ | $y \nVdash B$ |
| $a \subseteq b$ | $x \Vdash B$ | $y \Vdash C$ |
| $y \in b$ | $x \Vdash C$ | $z \Vdash A$ |
| $z \in a$ | $y \nVdash A$ | $z \Vdash C$ |




As we can see, $\forall\phi \in I_i(x)(\phi \Vdash^{\forall} C)$, so $x \Vdash K_i(C)$, there is an $a \in I_i(x)$ such that $a \Vdash^{\exists} C$ and $a \Vdash^{\forall} C \supset A$, so $x \Vdash_i A|C$, and moreover $x \Vdash \neg A$ and $x \Vdash B$. But there is a $b \in I_i(x)$ such that $b \nVdash^{\forall} A \vee B$, so $x \nVdash K_i(A \vee B)$. We are able to construct a countermodel because we can override the justification for Smith's belief (that $A|C$) with additional evidence. A clear contrast is visible with the branch ending in (3), which does not generate a counterexample, since such a situation cannot occur.

Notice that the piece of good epistemic luck ($x \Vdash B$) is insufficient for knowledge (since the example clearly states Smith has no reason to believe $B$ there is no $a \in I_i(x)$ such that $a \Vdash^{\forall} B$). Moreover, even if Smith had not had bad epistemic luck ($x \Vdash \neg A$) and the justification was warranted in $x$ ($x \in a$), it still *could* be overturned with additional evidence.

If we were to place $x$ anywhere else in the picture (e.g. within $a$ or $[\![A]\!] \cap [\![B]\!]$, or outside $[\![A]\!] \cup [\![B]\!]$) we would still have a counterexample to the inference of $K_i(A \vee B)$ from $K_i(C)$ and $x \Vdash_i A|C$. Smith does not have knowledge despite good luck because he cannot epistemically distinguish his situation from the one where his luck does not hold.

## 12.7 Stability Under Revision: Safe Belief and Knowledge

Now we can go back to the original question, i.e. what additional properties should be added to belief to turn it into knowledge. Everyday experience shows that beliefs can be shaken when some information, that may cause a change in beliefs, is acquired.

**Table 12.3** Rules for $Bel_i^{\mathrm{Sf}}$

**Rules for safe belief**

$$\frac{x \in a, a \in I_i(x), a \Vdash^{\forall} A, \psi \Rightarrow \Delta}{x : Bel_i^{\mathrm{Sf}} A, \psi \Rightarrow \Delta} \; LBel_i^{\mathrm{Sf}} \; (a \; fresh)$$

$$\frac{a \in I_i(x), x \in a, \psi \Rightarrow \Delta, x : Bel_i^{\mathrm{Sf}} A, a \Vdash^{\forall} A}{a \in I_i(x), x \in a, \psi \Rightarrow \Delta, x : Bel_i^{\mathrm{Sf}} A} \; RBel_i^{\mathrm{Sf}}$$

Beliefs can however be of varying strengths, depending on their degree of indefeasibility. We can require that a belief doesn't change whenever some true information is acquired. For example, a person believes indefeasibly in her partner's fidelity if she keeps believing it even when she reads authentic love letters written by her partner to a third party. We can further require the more extreme assumption that a belief is so strong that it cannot be questioned, even in the presence of misleading, false information, for example fabricated love letters or a planted handkerchief. The first type of belief is called *safe belief* and corresponds to "Stalnaker's knowledge" (Stalnaker, 1996), characterized as the doxastic attitude which is stable under revision with arbitrary *true* information.

In terms of epistemic plausibility models, the safe belief operator is defined as follows (Baltag & Smets, 2008; Pacuit , 2013):

$$(Safe_P) \quad \mathcal{M}_P, x \Vdash Bel_i^{\mathrm{Sf}} A \; iff \; \forall y \, . \, y \preceq_i x \; \rightarrow \; \mathcal{M}_P, y \Vdash A$$

and in terms of neighbourhood models (Girlando et al., 2018b):

$$(Safe_N) \quad \mathcal{M}_N, x \Vdash Bel_i^{\mathrm{Sf}} A \; iff \; \exists \phi \in I_i(x) \, . \, x \in \phi \; and \; \phi \Vdash^{\forall} A$$

We again follow the latter approach in adopting the rules for safe belief from Girlando et al. (2018b) (Table 12.3).

### *12.7.1 Properties of Safe Belief*

We first briefly show that all the usual structural properties hold for the rules for safe belief. As a first step we need to extend Definition 12.5.2 with the new case:

$$\mathrm{w}(Bel_i^{\mathrm{Sf}} A) = \mathrm{w}(A) + 1$$

This is sufficient to show that:

**Theorem 12.7.1** *For any $A$, the sequent $Bel_i^{Sf}A, \psi \Rightarrow \Delta, Bel_i^{Sf}A$ is derivable. Moreover, substitution, weakening and contraction are height-preserving admissible, and the rules $LBel_i^{Sf}$ and $RBel_i^{Sf}$ are height-preserving invertible. Finally, cut is admissible.*

***Proof*** The proofs are routine, so we just illustrate for cut, again for the case where the cut formula is principal in both premises, and of the form $Bel_i^{\mathrm{Sf}}A$:

$$\dfrac{\dfrac{x \in a, a \in I_i(x), \psi \Rightarrow \Delta, x : Bel_i^{\mathrm{Sf}}A, a \Vdash^{\forall} A}{x \in a, a \in I_i(x), \psi \Rightarrow \Delta, x : Bel_i^{\mathrm{Sf}}A}\, RBel_i^{\mathrm{Sf}} \qquad \dfrac{x \in b, b \in I_i(x), b \Vdash^{\forall} A, \psi' \Rightarrow \Delta'}{x : Bel_i^{\mathrm{Sf}}A, \psi' \Rightarrow \Delta'}\, LBel_i^{\mathrm{Sf}}}{x \in a, a \in I_i(x), \psi', \psi \Rightarrow \Delta, \Delta'}\, Cut$$

This is transformed into:

$$\dfrac{\dfrac{x \in a, a \in I_i(x), \psi \Rightarrow \Delta, x : Bel_i^{\mathrm{Sf}}A, a \Vdash^{\forall} A \qquad \dfrac{x \in b, b \in I_i(x), b \Vdash^{\forall} A, \psi' \Rightarrow \Delta'}{x : Bel_i^{\mathrm{Sf}}A, \psi' \Rightarrow \Delta'}\, {\scriptstyle LBel_i^{\mathrm{Sf}}}}{x \in a, a \in I_i(x), \psi', \psi \Rightarrow \Delta, \Delta', a \Vdash^{\forall} A}\, Cut_1 \qquad \dfrac{x \in b, b \in I_i(x), b \Vdash^{\forall} A, \psi' \Rightarrow \Delta'}{x \in a, a \in I_i(x), a \Vdash^{\forall} A, \psi' \Rightarrow \Delta'}\, Prop\ 12.5.1}{x \in a, x \in a, a \in I_i(x), a \in I_i(x), \psi', \psi' \Rightarrow \Delta, \Delta', \Delta'}\, Cut_2$$
$$x \in a, a \in I_i(x), \psi', \psi \Rightarrow \Delta, \Delta'\ \ Lm\ 12.5.6$$

where $\mathrm{Cut}_1$ is of lesser height and $\mathrm{Cut}_2$ is of lesser weight (recall the lexicographical ordering). □

Next we show that $Bel_i^{\mathrm{Sf}}$ is an S4.2 operator (Stalnaker, 2006):

**Theorem 12.7.2** *$Bel_i^{Sf}$ is an S4.2 operator, i.e. it satisfies the following axioms:*
(i) $Bel_i^{Sf}A \supset A$
(ii) $Bel_i^{Sf}A \supset Bel_i^{Sf}Bel_i^{Sf}A$
(iii) $\neg Bel_i^{Sf}\neg Bel_i^{Sf}A \supset Bel_i^{Sf}\neg Bel_i^{Sf}\neg A$

***Proof***

(i)

$$\dfrac{\dfrac{x \in a, a \in I_i(x), a \Vdash^{\forall} A, x : A \Rightarrow x : A}{x \in a, a \in I_i(x), a \Vdash^{\forall} A \Rightarrow x : A}\, L \Vdash^{\forall}}{x : Bel_i^{\mathrm{Sf}}A \Rightarrow x : A}\, LBel_i^{\mathrm{Sf}}$$

(ii)

$$\dfrac{\dfrac{\dfrac{\dfrac{\dfrac{\dfrac{\vdots}{a \in I_i(y), y \in a, x \in a, a \in I_i(x), a \Vdash^{\forall} A \Rightarrow x\colon Bel_i^{\mathrm{Sf}} Bel_i^{\mathrm{Sf}} A, y\colon Bel_i^{\mathrm{Sf}} A, a \Vdash^{\forall} A}\, RBel_i^{\mathrm{Sf}}}{a \in I_i(y), y \in a, x \in a, a \in I_i(x), a \Vdash^{\forall} A \Rightarrow x\colon Bel_i^{\mathrm{Sf}} Bel_i^{\mathrm{Sf}} A, y\colon Bel_i^{\mathrm{Sf}} A}\, A_1^*}{y \in a, x \in a, a \in I_i(x), a \Vdash^{\forall} A \Rightarrow x\colon Bel_i^{\mathrm{Sf}} Bel_i^{\mathrm{Sf}} A, y\colon Bel_i^{\mathrm{Sf}} A}\, R\Vdash^{\forall}}{x \in a, a \in I_i(x), a \Vdash^{\forall} A \Rightarrow x\colon Bel_i^{\mathrm{Sf}} Bel_i^{\mathrm{Sf}} A, a \Vdash^{\forall} Bel_i^{\mathrm{Sf}} A}\, RBel_i^{\mathrm{Sf}}}{x \in a, a \in I_i(x), a \Vdash^{\forall} A \Rightarrow x\colon Bel_i^{\mathrm{Sf}} Bel_i^{\mathrm{Sf}} A}\, LBel_i^{\mathrm{Sf}}}{x\colon Bel_i^{\mathrm{Sf}} A \Rightarrow x\colon Bel_i^{\mathrm{Sf}} Bel_i^{\mathrm{Sf}} A}$$

(iii)

$$\dfrac{\dfrac{\dfrac{\dfrac{\dfrac{\dfrac{

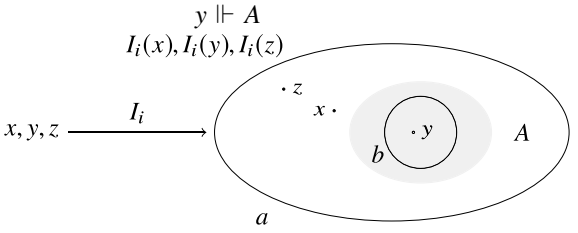
$$
\dfrac{\dfrac{\dfrac{\dfrac{\dfrac{\dfrac{\dfrac{\dfrac{\dfrac{\dfrac{\dfrac{y\colon A, b \subseteq a, b \in I_i(x), z \in a, y \in b, b \in I_i(y), b \Vdash^{\forall} A, y \in a, x \in a, a \in I_i(x) \Rightarrow ..., z\colon A}{b \subseteq a, b \in I_i(x), z \in a, y \in b, b \in I_i(y), b \Vdash^{\forall} A, y \in a, x \in a, a \in I_i(x) \Rightarrow ..., z\colon A}\, L\Vdash^{\forall}}{b \in I_i(x), z \in a, y \in b, b \in I_i(y), b \Vdash^{\forall} A, y \in a, x \in a, a \in I_i(x) \Rightarrow ..., z\colon A}\, S}{z \in a, y \in b, b \in I_i(y), b \Vdash^{\forall} A, y \in a, x \in a, a \in I_i(x) \Rightarrow ..., z\colon A}\, A_2}{z \in a, y\colon Bel_i^{\mathrm{Sf}} A, y \in a, x \in a, a \in I_i(x) \Rightarrow ..., z\colon A}\, LBel_i^{\mathrm{Sf}}}{y\colon Bel_i^{\mathrm{Sf}} A, y \in a, x \in a, a \in I_i(x) \Rightarrow ..., a \Vdash^{\forall} A}\, R\Vdash^{\forall}}{y \in a, x \in a, a \in I_i(x) \Rightarrow ..., y\colon \neg Bel_i^{\mathrm{Sf}} A, a \Vdash^{\forall} A}\, R\neg}{x \in a, a \in I_i(x) \Rightarrow ..., a \Vdash^{\forall} \neg Bel_i^{\mathrm{Sf}} A, a \Vdash^{\forall} A}\, R\Vdash^{\forall}}{x \in a, a \in I_i(x) \Rightarrow ..., a \Vdash^{\forall} \neg Bel_i^{\mathrm{Sf}} A, x\colon Bel_i^{\mathrm{Sf}} A}\, RBel_i^{\mathrm{Sf}}}{x \in a, a \in I_i(x) \Rightarrow x\colon Bel_i^{\mathrm{Sf}} \neg Bel_i^{\mathrm{Sf}} A, x\colon Bel_i^{\mathrm{Sf}} A}\, RBel_i^{\mathrm{Sf}}}{\Rightarrow x\colon Bel_i^{\mathrm{Sf}} \neg Bel_i^{\mathrm{Sf}} A, x\colon Bel_i^{\mathrm{Sf}} A}\, T}{x\colon \neg Bel_i^{\mathrm{Sf}} A \Rightarrow x\colon Bel_i^{\mathrm{Sf}} \neg Bel_i^{\mathrm{Sf}} A}\, L\neg
\quad (1)
$$

We could continue the derivation with several more applications of the $A$ rules, but we omit those for brevity. What those ultimately yield is $I_i(x) = I_i(y) = I_i(z)$, which we will have in the model. The proof search then stops at the sequent

$$
y\colon A, b \subseteq a, b \in I_i(x), z \in a, y \in b, b \in I_i(y), b \Vdash^{\forall} A, y \in a, x \in a, a \in I_i(x) \Rightarrow x\colon Bel_i^{\mathrm{Sf}} \neg Bel_i^{\mathrm{Sf}} A, x\colon Bel_i^{\mathrm{Sf}} A, z\colon A
$$

and we use it to construct a countermodel by assigning the value true to all the atomic formulas in the antecedent (including all relational atoms and the omitted $I_i(x) = I_i(y) = I_i(z)$) and false to all other atomic formulas, including relational atoms.

$I_i(x) = \{a, b\}$
$I_i(x) = I_i(y) = I_i(z)$
$b \subseteq a$
$x \in a$
$y \in a, b$
$z \in a$
$z \nVdash A$




Since $z \in a$ and $z \nVdash A$, then $\neg\forall y \in a.y \Vdash A$ and so $a \nVdash^{\forall} A$, thus $\neg\exists\phi \in I_i(x)$ $(x \in \phi \ \wedge \ \phi \Vdash^{\forall} A)$ and therefore $x \nVdash Bel_i^{\mathrm{Sf}} A$. So finally, $x \Vdash \neg Bel_i^{\mathrm{Sf}} A$.

Since $b = \{y\}$ and $y \Vdash A$ then $b \Vdash^{\forall} A$, and given that $y \in b$, $b \in I_i(y)$ it follows that $\exists\beta \in I_i(y)(y \in \beta \ \wedge \ \beta \Vdash^{\forall} A)$, so $y \Vdash Bel_i^{\mathrm{Sf}} A$ and therefore $y \nVdash \neg Bel_i^{\mathrm{Sf}} A$. Since $y \in a$, it follows that $a \nVdash^{\forall} \neg Bel_i^{\mathrm{Sf}} A$, thus $\neg\exists\phi \in I_i(x)(x \in \phi \ \wedge \ \phi \Vdash^{\forall} \neg Bel_i^{\mathrm{Sf}} A)$

(given that, since $x \notin b$, it holds by nesting for any $c \in I_i(x)$ that if $x \in c$ then $b \subseteq c$ and therefore $y \in c$), and so finally $x \nVdash Bel_i^{\mathrm{Sf}} \neg Bel_i^{\mathrm{Sf}} A$. □

### 12.7.2 Stability of Safe Belief

We now show that Stalnaker's knowledge corresponds to belief stable under arbitrary true revisions. In formal terms, this means that $Bel_i^{\mathrm{Sf}} A$ is true if and only if all conditional beliefs $Bel_i(A|P)$ are true, where $P$ is an arbitrary true proposition, that is, $Bel_i^{\mathrm{Sf}} A$ is equivalent to a meta-linguistic universal quantification, where $P$ ranges over true propositions.

We first show the more involved direction of the equivalence, that stability under true revision gives safe belief. For this, out of any true belief available for revision, we assume we can choose one (corresponding to an application of $L\forall$), which we designate $P^x$, such that it is "ordinary" for $i$ in relation to $x$, meaning that (i) it is credible for $i$, i.e. in some element of the neighbourhood $a \in I_i(x)$ it is universally forced, $a \Vdash^{\forall} P^x$ and (ii) plausible in $x$, i.e. true in neighbourhoods containing $x$ (and only there) , $b \Vdash^{\exists} P^x \leftrightarrow x \in b$. Notice that $P^x$ is true at $x$. For legibility we break down the proof into several sub-derivations (proof of *modus ponens* at the top of each derivation again omitted):

(1)

$$\dfrac{\dfrac{\dfrac{\dfrac{\dfrac{\dfrac{}{y \in b, y \in a, a \subseteq b, x \in b, b \in I(x), b \Vdash^{\exists} P^x, y\colon P^x \supset A, ..., y\colon P^x, ... \Rightarrow ..., y\colon A}\,\vdots}{y \in b, y \in a, a \subseteq b, x \in b, b \in I(x), b \Vdash^{\exists} P^x, y\colon P^x \supset A, ..., a \Vdash^{\forall} P^x, ... \Rightarrow ..., y\colon A}\, L\Vdash^{\forall}}{y \in b, y \in a, a \subseteq b, x \in b, b \in I(x), b \Vdash^{\exists} P^x, b \Vdash^{\forall} P^x \supset A, ..., a \Vdash^{\forall} P^x, ... \Rightarrow ..., y\colon A}\, L\Vdash^{\forall}}{y \in a, a \subseteq b, x \in b, b \in I(x), b \Vdash^{\exists} P^x, b \Vdash^{\forall} P^x \supset A, ..., a \Vdash^{\forall} P^x, ... \Rightarrow ..., y\colon A}\, L\subseteq}{a \subseteq b, x \in b, b \in I(x), b \Vdash^{\exists} P^x, b \Vdash^{\forall} P^x \supset A, a \in I(x), x \in a, a \Vdash^{\forall} P^x, ... \Rightarrow ..., a \Vdash^{\forall} A}\, R\Vdash^{\forall}}{a \subseteq b, x \in b, b \in I(x), b \Vdash^{\exists} P^x, b \Vdash^{\forall} P^x \supset A, a \in I(x), x \in a, a \Vdash^{\forall} P^x, ... \Rightarrow x\colon Bel_i^{\mathrm{Sf}}(A)}\, RBel_i^{\mathrm{Sf}}$$

(2)

$$\dfrac{\dfrac{\dfrac{\dfrac{\dfrac{\dfrac{}{y \in a, y \in b, b \subseteq a, x \in b, b \in I(x), b \Vdash^{\exists} P^x, y\colon P^x \supset A, ..., y\colon P^x, ... \Rightarrow ..., y\colon A}\,\vdots}{y \in a, y \in b, b \subseteq a, x \in b, b \in I(x), b \Vdash^{\exists} P^x, y\colon P^x \supset A, ..., a \Vdash^{\forall} P^x, ... \Rightarrow ..., y\colon A}\, L\Vdash^{\forall}}{y \in a, y \in b, b \subseteq a, x \in b, b \in I(x), b \Vdash^{\exists} P^x, b \Vdash^{\forall} P^x \supset A, ..., a \Vdash^{\forall} P^x, ... \Rightarrow ..., y\colon A}\, L\Vdash^{\forall}}{y \in b, b \subseteq a, x \in b, b \in I(x), b \Vdash^{\exists} P^x, b \Vdash^{\forall} P^x \supset A, ..., a \Vdash^{\forall} P^x, ... \Rightarrow ..., y\colon A}\, L\subseteq}{b \subseteq a, x \in b, b \in I(x), b \Vdash^{\exists} P^x, b \Vdash^{\forall} P^x \supset A, a \in I(x), x \in a, a \Vdash^{\forall} P^x, ... \Rightarrow ..., b \Vdash^{\forall} A}\, R\Vdash^{\forall}}{b \subseteq a, x \in b, b \in I(x), b \Vdash^{\exists} P^x, b \Vdash^{\forall} P^x \supset A, a \in I(x), x \in a, a \Vdash^{\forall} P^x, ... \Rightarrow x\colon Bel_i^{\mathrm{Sf}}(A)}\, RBel_i^{\mathrm{Sf}}$$

(3)

$$\dfrac{\dfrac{(1) \qquad (2)}{x \in b, b \in I(x), b \Vdash^{\exists} P^x, b \Vdash^{\forall} P^x \supset A, a \in I(x), x \in a, a \Vdash^{\forall} P^x, x : Bel_i(A|P^x) \Rightarrow x : Bel_i^{\mathrm{Sf}}(A)}\, S}{b \in I(x), b \Vdash^{\exists} P^x, b \Vdash^{\forall} P^x \supset A, a \in I(x), x \in a, a \Vdash^{\forall} P^x, x : Bel_i(A|P^x) \Rightarrow x : Bel_i^{\mathrm{Sf}}(A)}\, P^x$$
$$\dfrac{}{x \Vdash_i A|P^x, a \in I(x), x \in a, a \Vdash^{\forall} P^x, x : Bel_i(A|P^x) \Rightarrow x : Bel_i^{\mathrm{Sf}}(A)}\, LC$$

$$\dfrac{\dfrac{\dfrac{\dfrac{\dots, x \in a, a \Vdash^{\forall} P^x, x : P^x, \dots \Rightarrow \dots, a \Vdash^{\exists} P^x, x : P^x}{\dots, x \in a, a \Vdash^{\forall} P^x, x : P^x, \dots \Rightarrow \dots, a \Vdash^{\exists} P^x}\, L\Vdash^{\exists}}{\dots, x \in a, a \Vdash^{\forall} P^x, \dots \Rightarrow \dots, a \Vdash^{\exists} P^x}\, L\Vdash^{\forall}}{a \in I(x), x \in a, a \Vdash^{\forall} P^x, x : Bel_i(A|P^x) \Rightarrow x : Bel_i^{\mathrm{Sf}}(A)}\, (3)\, LB_i$$

Next, we demonstrate the other direction, that safe belief is stable under true revisions. Here no other requirement is needed except $P$ being true at $x$.

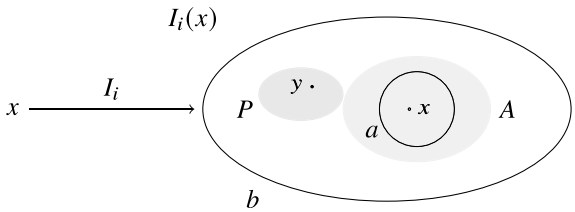
$$\dfrac{\dfrac{\dfrac{\vdots \qquad\qquad \vdots}{\dots x \in a, x : P \Rightarrow a \Vdash^{\exists} P \quad \dots x \in a, a \Vdash^{\forall} A, x : P, b \Vdash^{\exists} P \Rightarrow a \Vdash^{\forall} P \supset A}\, RC}{a \in I_i(x), b \in I_i(x), x \in a, a \Vdash^{\forall} A, x : P, b \Vdash^{\exists} P \Rightarrow x \Vdash_i A|P}\, RB_i}{a \in I_i(x), x \in a, a \Vdash^{\forall} A, x : P \Rightarrow x : Bel_i(A|P)}\, LBel_i^{\mathrm{Sf}}$$
$$x : Bel_i^{\mathrm{Sf}} A, x : P \Rightarrow x : Bel_i(A|P)$$

It is obvious that safe belief will not be stable under *arbitrary* revision—without the assumption that $P$ is true at $x$ the left branch stops at the sequent (after a few simple steps)

$$x \in b, a \subseteq b, a \in I_i(x), b \in I_i(x), x \in a, a \Vdash^{\forall} A, x : A, y \in B, y : P \Rightarrow a \Vdash^{\exists} P, x : P$$

and we obtain the following countermodel:

$I_i(x) = \{a, b\}$
$a \subseteq b$
$x \in a, b$
$y \in b$
$x \Vdash A, x \nVdash P$
$y \Vdash P, y \nVdash A$



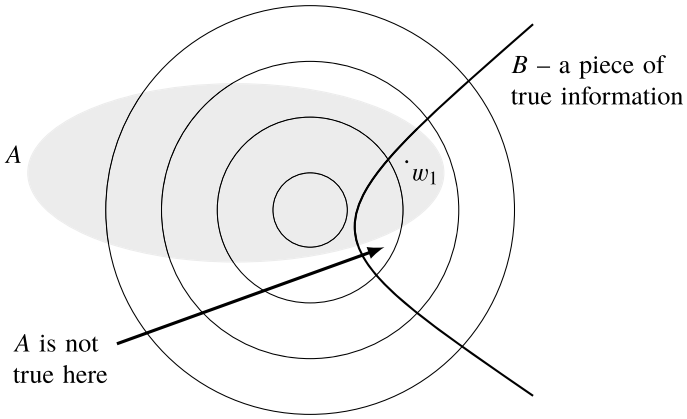

Since $x \Vdash A$ and $x$ is the only element of $a$ it follows that $a \Vdash^{\forall} A$, and from there, $a \in I_i(x)$ and $x \in a$ it follows that $x \Vdash_i Bel_i^{\mathrm{Sf}} A$.

Since $b \in I_i(x)$, $y \in b$ and $y \Vdash P$ it follows that $\neg(\forall \phi \in I_i(x).\phi \Vdash^{\forall} \neg P)$ (the first disjunct of Definition 12.3.7 doesn't hold). Since $x \nVdash P$ and $x$ is the only element of $a$ it follows that $a \nVdash_i^{\exists} P$. From $y \in b$, $y \Vdash P$ and $y \nVdash A$ it follows that $b \nVdash_i^{\forall} P \supset A$.

Given that $I_i(x) = \{a, b\}$, it follows that $\neg\exists\beta \in I_i(x).\beta \Vdash^{\exists} P$ *and* $\beta \Vdash^{\forall} P \supset A$ (the second disjunct of Definition 12.3.7 doesn't hold). Therefore finally $x \nVdash Bel_i(A|P)$.

### *12.7.3 Stability of Knowledge*

We now show that, in contrast, knowledge does have the stronger property of stability under revision with *arbitrary* (including deceitful) information. We first present a graphic illustration of the idea of this proof from Rott (2004) (as in the other illustrations in this paper, the concentric circles represent neighbourhoods and the gray area stands for $[\![A]\!]$, but here $[\![B]\!]$ is the area to the right of the parabola):




The picture illustrates that if we deny knowledge of $A$, then it is possible to find a proposition $B$ that destroys conditional belief in $A$. An additional assumption required for the graphic proof to work is non-emptiness—for all $b$ such that $b \Vdash^{\exists} B$ we have $y \in b$ ($y$ is the point indicated by the arrow).

We now present the sequent calculus proof. In the first direction, we show that belief stable under arbitrary revisions is knowledge. We choose one particular revision (a deceitful one!), corresponding to a universal instantiation (i.e. L$\forall$) of a metatheoretical fact:

(1)

$$\dfrac{\dfrac{\vdots}{y \in a, a \in I_i(x), x : Bel_i(A|\neg A) \Rightarrow y : A, a \Vdash^{\exists} \neg A, y : \neg A}\, R\Vdash^{\exists}}{a \in I_i(x), x : Bel_i(A|\neg A) \Rightarrow a \Vdash^{\forall} A, a \Vdash^{\exists} \neg A}\, R\Vdash^{\forall}$$

(2)

$$\dfrac{\dfrac{\dfrac{\dfrac{\vdots}{z \in b, z\colon \neg A, z\colon \neg A \supset A, b \in I_i(x), b \Vdash^{\forall} \neg A \supset A, y \in a, a \in I_i(x), x\colon Bel_i(A|\neg A) \Rightarrow y\colon A}\, L\Vdash^{\forall}}{z \in b, z\colon \neg A, b \in I_i(x), b \Vdash^{\forall} \neg A \supset A, y \in a, a \in I_i(x), x\colon Bel_i(A|\neg A) \Rightarrow y\colon A}\, L\Vdash^{\exists}}{b \in I_i(x), b \Vdash^{\exists} \neg A, b \Vdash^{\forall} \neg A \supset A, y \in a, a \in I_i(x), x\colon Bel_i(A|\neg A) \Rightarrow y\colon A}\, LC}{x \Vdash_i A|\neg A, y \in a, a \in I_i(x), x\colon Bel_i(A|\neg A) \Rightarrow y\colon A}$$

$$\dfrac{\dfrac{(1) \qquad (2)}{a \in I_i(x), x\colon Bel_i(A|\neg A) \Rightarrow a \Vdash^{\forall} A}\, LB_i}{x\colon Bel_i(A|\neg A) \Rightarrow x\colon K_i A}\, RK_i$$

The other direction is more straightforward. Here we prove that knowledge is stable under arbitrary revisions, corresponding to universal generalization (i.e. R∀) to a metatheoretical fact, and thus our choice of a revision must be completely arbitrary:

$$\dfrac{\dfrac{\dfrac{\dfrac{\dfrac{\vdots}{\ldots a \Vdash^{\exists} P, x\colon K_i A \Rightarrow a \Vdash^{\exists} P} \quad \dfrac{\vdots}{a \in I_i(x), a \Vdash^{\forall} A, a \Vdash^{\exists} P, x\colon K_i A \Rightarrow a \Vdash^{\forall} P \supset A}\, RC}{a \in I_i(x), a \Vdash^{\forall} A, a \Vdash^{\exists} P, x\colon K_i A \Rightarrow x \Vdash_i A|P}\, LK_i}{a \in I_i(x), a \Vdash^{\exists} P, x\colon K_i A \Rightarrow x \Vdash_i A|P}\, RB_i}{x\colon K_i A \Rightarrow x\colon Bel_i(A|P)}$$

The sequent calculus proof is the formalization of a model-theoretic proof by contradiction (the negative assumption corresponding to the occurrence of $x\colon K_i A$ in the succedent), however it is a direct proof, since no explicit use of indirect arguments is needed because of the classical base of the calculus.

## 12.8 Proof Theory and Paradox Control

As already witnessed in earlier work, cf. Maffezioli et al. (2013) for the Fitch paradox of knowability, a proof-theoretical approach to formal epistemology gives a fine-grained method to first analyze paradoxes in order to find the source of a paradoxical conclusion and then to propose solutions.

In more specific terms, a ground sequent calculus with invertible rules gives, through proof search for specific axioms, a heuristic for finding the corresponding semantic properties and the extension to a complete calculus for the particular logic in question. Analytic proof systems give control not only of provability but also of *unprovability*, thus giving a definite argument for paradox defeat and definite answers to revenge arguments.

One could argue that the use of purely semantic methods does succeed in producing countermodels, and thus likewise makes conclusions about underivability. We notice, however, that such methods are not fully independent, as they rest on completeness results, nor fully algorithmic. As emphasized in the introduction, labelled sequent calculi include, in fact, the strength of the semantic methods, but make it in-built in a two-fold mechanism of invertible rules: invertibility is used in a bootstrap fashion at the level of the genesis of calculi for finding the rules needed, and at the object level for finding derivations (or countermodels) in a guided way. Pinpointing the necessary rules, also gives indication for paradox defeat. This is best seen in concrete terms in what follows.

### 12.8.1 Argument Against a Perfect Believer

The paradox of the perfect believer is a derivation of an *implication from belief of knowledge to knowledge* from apparently reasonable assumptions on the classical epistemic/doxastic operators (Gochet & Gribomont, 2006), which in addition to axioms (K), (T) and (5) for knowledge and (K) for belief include: (i) weak infallibility, $\neg Bel_i \bot$, (ii) that knowledge implies belief, $K_i A \supset Bel_i A$, and (iii) introspection about belief $Bel_i A \supset K_i Bel_i A$.

We now prove that the assumptions used in the puzzle hold.
Weak infallibility, $\neg Bel_i\ \bot$, follows from $N$ (*non-emptiness*):

$$\dfrac{\dfrac{\dfrac{\dfrac{\dfrac{}{y : \bot, y \in a, a \in I(x), a \Vdash^{\forall} \bot \Rightarrow}\, L\bot}{y \in a, a \in I(x), a \Vdash^{\forall} \bot \Rightarrow}\, L\Vdash^{\forall}}{a \in I(x), a \Vdash^{\forall} \bot \Rightarrow}\, N}{x : Bel_i \bot \Rightarrow}\, LSB_i}{\Rightarrow x : \neg Bel_i \bot}\, R\neg$$

$K_i A \supset Bel_i A$ is valid thanks to $T$ (*total reflexivity*):

$$\dfrac{\dfrac{\dfrac{\dfrac{y \in a, x \in a, a \in I(x), x : K_i A, a \Vdash^{\forall} A, y : A \Rightarrow x : Bel_i A, y : A}{y \in a, x \in a, a \in I(x), x : K_i A, a \Vdash^{\forall} A \Rightarrow x : Bel_i A, y : A}\, L\Vdash^{\forall}}{x \in a, a \in I(x), x : K_i A, a \Vdash^{\forall} A \Rightarrow x : Bel_i A, a \Vdash^{\forall} A}\, R\Vdash^{\forall}}{x \in a, a \in I(x), x : K_i A, a \Vdash^{\forall} A \Rightarrow x : Bel_i A}\, RSB_i}{x \in a, a \in I(x), x : K_i A \Rightarrow x : Bel_i A}\, LK_i}{x : K_i A \Rightarrow x : Bel_i A}\, T$$

$Bel_i A \supset K_i Bel_i A$ is valid thanks to $A$ rules (*local absoluteness*):

$$\dfrac{\dfrac{\dfrac{\dfrac{\dfrac{\dfrac{\dfrac{z \in a, a \in I(y), y \in b, b \in I(x), a \in I(x), a \Vdash^{\forall} A, z\colon A \Rightarrow y\colon Bel_i A, z\colon A}{z \in a, a \in I(y), y \in b, b \in I(x), a \in I(x), a \Vdash^{\forall} A \Rightarrow y\colon Bel_i A, z\colon A}\, L\Vdash^{\forall}}{a \in I(y), y \in b, b \in I(x), a \in I(x), a \Vdash^{\forall} A \Rightarrow y\colon Bel_i A, a \Vdash^{\forall} A}\, R\Vdash^{\forall}}{a \in I(y), y \in b, b \in I(x), a \in I(x), a \Vdash^{\forall} A \Rightarrow y\colon Bel_i A}\, RSB_i}{y \in b, b \in I(x), a \in I(x), a \Vdash^{\forall} A \Rightarrow y\colon Bel_i A}\, A_1}{b \in I(x), a \in I(x), a \Vdash^{\forall} A \Rightarrow b \Vdash^{\forall} Bel_i A}\, R\Vdash^{\forall}}{a \in I(x), a \Vdash^{\forall} A \Rightarrow x\colon K_i Bel_i A}\, RK_i}{x\colon Bel_i A \Rightarrow x\colon K_i Bel_i A}\, LSB_i$$

The derivation of the paradox is as follows:

(1)

$$\dfrac{\dfrac{\dfrac{\dfrac{\dfrac{\dfrac{\dfrac{\dfrac{\dfrac{\dfrac{\dfrac{\dots, y\colon A \Rightarrow y\colon A}{b \in I_i(z), \dots, \dots, y \in b, \dots, b \Vdash^{\forall} A, \dots \Rightarrow y\colon A}\, L\Vdash^{\forall}}{b \in I_i(z), \dots, \dots, y \in b, \dots, z\colon K_i A, \dots \Rightarrow y\colon A}\, LK_i}{z \in c, c \in I_i(z), z \in a, z \in b, a \subseteq b, a \in I_i(x), b \in I_i(x), y \in b, a \Vdash^{\forall} K_i A, z\colon K_i A, c \Vdash^{\forall} A, z\colon A \Rightarrow y\colon A}\, A_1}{z \in c, c \in I_i(z), z \in a, z \in b, a \subseteq b, a \in I_i(x), b \in I_i(x), y \in b, a \Vdash^{\forall} K_i A, z\colon K_i A, c \Vdash^{\forall} A \Rightarrow y\colon A}\, L\Vdash^{\forall}}{z \in c, c \in I_i(z), z \in a, z \in b, a \subseteq b, a \in I_i(x), b \in I_i(x), y \in b, a \Vdash^{\forall} K_i A, z\colon K_i A \Rightarrow y\colon A}\, LK_i}{z \in a, z \in b, a \subseteq b, a \in I_i(x), b \in I_i(x), y \in b, a \Vdash^{\forall} K_i A, z\colon K_i A \Rightarrow y\colon A}\, T}{z \in a, z \in b, a \subseteq b, a \in I_i(x), b \in I_i(x), y \in b, a \Vdash^{\forall} K_i A \Rightarrow y\colon A}\, L\Vdash^{\forall}}{z \in a, a \subseteq b, a \in I_i(x), b \in I_i(x), y \in b, a \Vdash^{\forall} K_i A \Rightarrow y\colon A}\, L\subseteq}{a \subseteq b, a \in I_i(x), b \in I_i(x), y \in b, a \Vdash^{\forall} K_i A \Rightarrow y\colon A}\, N$$

$$\dfrac{\dfrac{\dfrac{\dfrac{\dfrac{\dfrac{\dfrac{\dfrac{\dots, y\colon A \Rightarrow y\colon A}{\dots, y \in d, d \in I_i(y), d \Vdash^{\forall} A \Rightarrow y\colon A}\, L\Vdash^{\forall}}{\dots, y \in d, d \in I_i(y), y\colon K_i A \Rightarrow y\colon A}\, LK_i}{\dots, y\colon K_i A \Rightarrow y\colon A}\, T}{y \in a, \dots a \Vdash^{\forall} K_i A \Rightarrow y\colon A}\, L\Vdash^{\forall}}{b \subseteq a, \dots, y \in b, a \Vdash^{\forall} K_i A \Rightarrow y\colon A}\, L\subseteq}{a \in I_i(x), b \in I_i(x), y \in b, a \Vdash^{\forall} K_i A \Rightarrow y\colon A}\, (1)\ S}{a \in I_i(x), b \in I(x)_i, a \Vdash^{\forall} K_i A \Rightarrow b \Vdash^{\forall} A}\, R\Vdash^{\forall}}{a \in I_i(x), a \Vdash^{\forall} K_i A \Rightarrow x\colon K_i A}\, RK_i}{x\colon Bel_i K_i A \Rightarrow x\colon K_i A}\, LSB_i$$

Given factivity of knowledge, this means we could not believe to know something false, thus making us perfect believers (Gochet & Gribomont, 2006). Moreover, combined with the strong understanding of belief as belief with certainty (Baltag and Smets 2016), $Bel_i A \supset Bel_i K_i A$, knowledge and belief collapse (we have already seen that $K_i A \supset Bel_i A$, so we just need the implication in the other direction):

$$\dfrac{\dfrac{\begin{array}{c}\vdots\\ x : Bel_i A \Rightarrow x : K_i A, x : Bel_i A \end{array} \quad x : Bel_i K_i A, x : Bel_i A \Rightarrow x : K_i A}{x : Bel_i A \supset Bel_i K_i A, x : Bel_i A \Rightarrow x : K_i A}\, L\supset}{x : Bel_i A \supset Bel_i K_i A \Rightarrow x : Bel_i A \supset K_i A}\, R\supset$$

We now examine various methods of resolving the paradox by omitting the prerequisite semantic conditions.

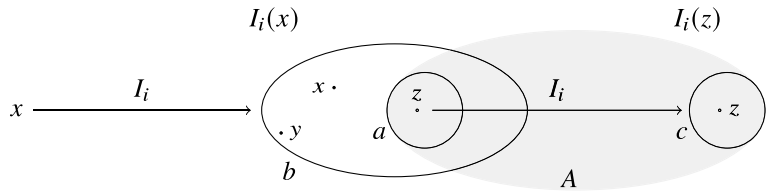
#### 12.8.1.1 Resolution 1: Local Absoluteness

The first method of blocking the derivation is by omitting the rule $A_1$, corresponding to the left-to-right direction of local absoluteness, $(a \in I_i(x) \wedge y \in a) \supset (b \in I_i(x) \supset b \in I_i(y))$. In that case the proof search stops at the sequent

$$z \in c, c \in I_i(z), z \in a, z \in b, a \subseteq b, a \in I_i(x), b \in I_i(x), y \in b,$$
$$a \Vdash^{\forall} K_i A, z : K_i A, c \Vdash^{\forall} A, z : A \Rightarrow y : A$$

and we obtain the countermodel from the failed proof search on the right-hand branch. This is achieved by assigning true to all the atomic formulas in the antecedent (including all relational atoms) and false to all other relational atoms and all the formulas in the succedent. In this way we obtain the following countermodel:

$$\left.\begin{array}{l} I_i(x) = \{a, b\} \\ I_i(z) = \{c\} \\ z \in a, b, c \end{array}\right\} \text{(only without } A_1\text{)} \qquad \begin{array}{l} y \notin a \\ a \subseteq b \\ z \Vdash A \\ y \nVdash A \end{array}$$
$$y \in b$$




Note that nesting is satisfied for both $I_i(x)$ and $I_i(z)$. Moreover, we simply place $x$ somewhere to satisfy total reflexivity (instead of making the picture more complicated by adding another neighbourhood), but as will be seen below its location does not affect the counterexample. Finally, non-emptiness is satisfied.

Since $y \nVdash A$ and $y \in b$ then $b \nVdash^{\forall} A$, and since $b \in I_i(x)$ it follows that $\neg\forall\beta \in I_i(x)$ $(\beta \Vdash^{\forall} A)$, and therefore $x \nVdash K_i A$. Since $z \Vdash A$ and $z$ is the only element of $c$ (and in any case for every $u \in c$ it would hold that $u \Vdash A$ by L$\Vdash^{\forall}$) it follows that

$c \Vdash^{\forall} A$. Since $I_i(z) = \{c\}$, it holds that $\forall \beta \in I_i(z)\ (\beta \Vdash^{\forall} A)$, therefore $z \Vdash K_i A$ and thus $a \Vdash^{\forall} K_i A$. So, $\exists \phi \in I_i(x)\ (\phi \Vdash^{\forall} K_i A)$ and finally $\Vdash Bel_i K_i A$.

Recall that $A_1$ was required for the derivation of negative introspection (Theorem 12.6.6.iii) and so this shows us that, if we reject negative introspection, the paradox dissipates. Given that it is generally a controversial property (Williamson, 2002), this is a common response to the paradox; see e.g. van der Hoek (1993) and Williamson (2001).

Also akin to this solution is one option proposed by Baltag and Smets (2016), where knowledge is substituted for safe belief—recall that (Theorem 12.7.3) negative introspection does not hold for the latter. An argument in favour of S4.2 as logic of knowledge can also be found in (Lenzen, 1979).

This likewise blocks the collapse of knowledge and belief, even though belief-as-belief-with-certainty principle ($Bel_i A \supset Bel_i Bel_i^{\mathrm{Sf}} A$) holds for safe belief.

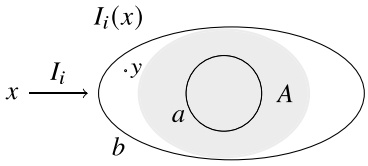
#### 12.8.1.2 Resolution 2: Non-emptiness

Observing the derivation of the paradox, another method of resolution now presents itself. We now notice that without $N$ proof search stops at the sequent

$$a \subseteq b, a \in I_i(x), b \in I_i(x), y \in b, a \Vdash^{\forall} K_i A \Rightarrow y : A$$

and we obtain the following countermodel:

$I_i(x) = \{a, b\}$
$a \subseteq b$
$y \in b$
$a = \emptyset$ (only possible without N)
$a \Vdash^{\forall} K_i A$
$y \nVdash A$




Since $a \Vdash^{\forall} K_i A$ (vacuously), then $\exists \phi \in I_i(x)(\phi \Vdash^{\forall} K_i A)$, and therefore $x \Vdash Bel_i K_i A$. But since $y \in b$, $y \nVdash A$, then $b \nVdash^{\forall} A$, and therefore $\neg \forall \beta \in I_i(x)(\beta \Vdash^{\forall} A)$, so $x \nVdash K_i A$.

Recall that $N$ was required for the derivation of weak infallibility (notice that in the countermodel $x \Vdash Bel_i \neg K_i A$ also holds), so abandoning that semantic condition likewise blocks one of the ingredients for the paradox. Though a much less common option, this method of resolution is nonetheless discussed by Gochet and Gribomont (2006).

### 12.8.2 *"I Know Only What I Believe"*

Luckily (or maybe not) the political scene has become a source of interesting epistemic puzzles; this particular sentence is famous from a speech by Tony Blair about the Iraqi war in 2004, part of the statement "I know the issue has divided the country. [...] I'm like any other human being as fallible and as capable of being wrong. I only know what I believe."[8] Of this passage, Timothy Williamson wrote:

> He hadn't known that there were WMD, but he had known that he believed that there were WMD. He tried to divert attention away from the question of testable evidence out there for WMD to the question of his inner sincerity. (Williamson, 2018, p. 2)

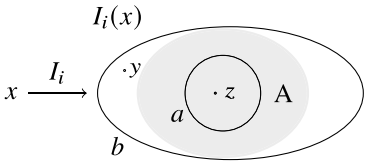
And in an interwiew:

> When it turned out that there were no weapons of mass destruction in Iraq (the original excuse for invading in 2003), the public debate was confused by failure to make the basic epistemological distinction between knowledge and belief. Tony Blair insisted that he had genuinely believed that there were WMD in Iraq, obscuring the key point that he hadn't known that there were. (Sosis, 2017)

If one takes this as substituting knowledge of facts by knowledge of personal beliefs, as in the first Williamson quote, then this is an instance of the perfect believer paradox. If we take it at face value, like in the second quote, then this conflation of knowledge and belief (of which the latter is fallible) is a simpler case, as we obtain almost the same countermodel as for the "I know only what I believe I know", and this works also in the presence of the assumption $N$ (for brevity we omit an application of $S$).

$$\dfrac{\dfrac{\dfrac{\dfrac{\dfrac{z \in a, y \in b, a \in I_i(x), b \in I_i(x), z\colon A \Rightarrow y\colon A}{z \in a, y \in b, a \in I_i(x), b \in I_i(x), a \Vdash^{\forall} A \Rightarrow y\colon A}\, L\Vdash^{\forall}}{y \in b, a \in I_i(x), b \in I_i(x), a \Vdash^{\forall} A \Rightarrow y\colon A}\, N}{a \in I_i(x), b \in I_i(x), a \Vdash^{\forall} A \Rightarrow b \Vdash^{\forall} A}\, R\Vdash^{\forall}}{a \in I_i(x), a \Vdash^{\forall} A \Rightarrow x\colon K_i A}\, RK_i}{x\colon Bel_i A \Rightarrow x\colon K_i A}\, LSB_i$$

$I_i(x) = \{a, b\}$
$a \subseteq b$
$y \in b\,; z \in a, b$
$y \nVdash A$
$z \Vdash A$




[8] For a full transcript of the speech, see Blair (2004) at: https://www.theguardian.com/politics/2004/sep/28/labourconference.labour5.

Our agent is not saved, however, even by strong fallibility (absence of $N$). Observe that we then also have a countermodel, with empty $a$ (the same countermodel from the previous section), so fallibility is not a reason that justifies the inference of knowledge from belief.

## 12.9 Conclusion

In this paper we have made a contribution to the proof theory of dynamic epistemic logic, which, as we have pointed out in the opening, is significantly less developed than its semantic counterpart. In particular, we explored and expanded on the logic of conditional belief CDL, from its presentation in (Girlando et al., 2018b), and in particular used it to address the problem of defining knowledge.

We recalled how the connection between more general neighbourhood models and relational (Kripke) models and plausibility models is established, and further developed on their basis complete proof systems following a methodology introduced by Negri (2017b) and developed for conditional doxastic notions by Girlando et al. (2018b).

The resulting labelled sequent calculus in which knowledge and (simple) belief are obtained as degenerate cases of conditional belief has then been shown to possess the standard array of desirable structural properties, including admissibility of weakening, contraction and cut.

Moreover, we have provided the rules for the safe belief operator and proved that it can be characterized as a doxastic attitude which is stable under revision with arbitrary true information. Knowledge defined in CDL has instead been shown to be stable under revision by arbitrary (including deceitful) information.

Finally, we have shown how our proof theory gives a powerful tool for paradox control by demonstrating this with several examples. Crucial here is that the use of a sequent calculus with invertible rules provides both a way to find the required rules to obtain specific systems, as well as formal proofs of underivability of certain properties, thus giving definite arguments for paradox defeat.

**Acknowledgements** This paper was presented in 2018 in Amsterdam at the conference "Reasoning in Social Context." The authors are grateful to the organizers and the other participants for insightful comments. The research was partially supported by the Academy of Finland, research project no. 1308664. The authors extend their gratitude to anonymous referees for their extensive and perspicacious suggestions which contributed greatly to the present state of the paper.

## References


Alchourrón, C. E., Gärdenfors, P., & Makinson, D. (1985). On the logic of theory change: Partial meet contraction and revision functions. *Journal of Symbolic Logic, 50*(2), 510–530.

Aucher, G., Maubert, B., & Schwarzentruber, F. (2012). Generalized DEL-Sequents Generalized DEL-sequents. In L. F. del Cerro, A. Herzig, & J. Mengin (Eds.), *Logics in artificial intelligence* (pp. 54–66). Berlin, Heidelberg: Springer.

Baltag, A. , & Renne, B. (2016). Dynamic epistemic logic . In E. N. Zalta (Ed.), *The Stanford encyclopedia of philosophy* (Winter 2016 Ed.). Metaphysics Research Lab, Stanford University. https://plato.stanford.edu/archives/win2016/entries/dynamic-epistemic/. [Accessed: 06/18/2021].

Baltag, A., & Smets, S. (2006). Conditional doxastic models: A qualitative approach to dynamic belief revision. *Electronic Notes in Theoretical Computer Science, 165*, 5–21.

Baltag, A., & Smets, S. (2008). The logic of conditional doxastic actions. In K. Apt & R. Van Rooij (Eds.), *New perspectives on games and interaction* (Vol. 4, pp. 9–31). Amsterdam: Amsterdam University Press.

Baltag, A., & Smets, S. (2016). A qualitative theory of dynamic interactive belief revision. In H. Arló-Costa, V. F. Hendricks, & J. van Benthem (Eds.), *Readings in formal epistemology: Sourcebook* (pp. 813–858). Cham: Springer International Publishing.

Basin, D., Matthews, S., & Viganó, L. (1998). Natural deduction for non-classical logics. *Studia Logica, 60*, 119–160.

Blackburn, P. (2000). Internalizing labelled deduction. *Journal of Logic and Computation, 10*(1), 137–168. https://doi.org/10.1093/logcom/10.1.137

Blair, T. (2004). Full text: Blair's conference speech (part two). In *The Guardian*. https://www.theguardian.com/politics/2004/sep/28/labourconference.labour5. [Accessed: 06/20/2021].

Board, O. (2004). Dynamic interactive epistemology. *Games and Economic Behavior, 49*(1), 49–80. https://doi.org/10.1016/j.geb.2003.10.006

Castellini, C. (2006). Automated reasoning in quantified modal and temporal logics. *AI Communications, 19*(2), 183–185.

Castellini, C., & Smaill, A. (2002). A systematic presentation of quantified modal logics. *Logic Journal of the IGPL, 10*(6), 571–599.

Catach, L. (1991). Tableaux: A general theorem prover for modal logics. *Journal of Automated Reasoning, 7*, 489–510.

Chellas, B. (1980). *Modal logic: An introduction*. Cambridge: Cambridge University Press.

del Cerro, L.F., Fauthoux, D., Gasquet, O., Herzig, A. , Longin, D., & Massacci, F. (2001). Lotrec: The generic tableau prover for modal and description logics. In *International joint conference on automated reasoning* (pp. 453–458). Berlin, Heidelberg.

Dyckhoff, R., & Sadrzadeh, M. (2010). Positive logic with adjoint modalities: Proof theory, semantics, and reasoning about information. *The Review of Symbolic Logic, 3*(3), 351–373. https://doi.org/10.1017/S1755020310000134

Fitch, F. (1966). Tree proofs in modal logic (abstract). *Journal of Symbolic Logic, 31*(1), 152.

Fitting, M. (1983). *Proof methods for modal and intuitionistic logic* (Vol. 169). D. Reidel Publishing Company, Dordrecht, Boston, and Lancaster.

Frittella, S., Greco, G., Kurz, A., Palmigiano, A., & Sikimić, V. (2016). Multi-type display calculus for dynamic epistemic logic. *Journal of Logic and Computation, 26*(6), 2017–2065. https://doi.org/10.1093/logcom/exu068.

Gabbay, D. (1996). *Labelled deductive systems*. Oxford: Oxford University Press.

Gettier, E. L. (1963). Is justified true belief knowledge? *Analysis, 23*(6), 121–123.

Girlando, M., Lellmann, B., & Olivetti, N. (2018a). Hypersequent calculus for the logic of conditional belief: Preliminary results. In *EICNCL Floc Workshop Proceedings*.

Girlando, M., Negri, S., Olivetti, N., & Risch, V. (2018b). Conditional beliefs: From neighbourhood semantics to sequent calculus. *The Review of Symbolic Logic, 11*(4), 736–779. https://doi.org/10.1017/S1755020318000023

Girlando, M., Negri, S., & Olivetti, N. (2021). Uniform labelled calculi for preferential conditional logics based on neighbourhood semantics. *Journal of Logic and Computation, 31*(3), 947–997.
Gochet, P., & Gribomont, P. (2006). Epistemic logic. In D. M. Gabbay & J. H. Woods (Eds.), *Handbook of the history of logic* (Vol. 7, pp. 99–195). Elsevier North-Holland.
Goré, R. (1999). Tableau methods for modal and temporal logics. In M. D'Agostino, D. Gabbay, R. Hähnle, & J. Posegga (Eds.), *Handbook of tableau methods* (pp. 297–396). Springer.
Grove, A. (1988). Two modellings for theory change. *Journal of Philosophical Logic, 17*(2), 157–170.
Hintikka, J. (1962). *Knowledge and belief: An introduction to the logic of the two notions*. Ithaca, NY: Cornell University Press.
Indrzejczak, A. (2010). *Natural deduction, hybrid systems and modal logics* (Vol. 30). Dordrecht, Heidelberg, London, New York: Springer Science & Business Media.
Jowett, B., & Pelliccia, H. (2001). *Selected dialogues of Plato: The Benjamin Jowett translation*. Modern Library.
Kanger, S. (1957). *Provability in logic*. Stockholm: Almqvist & Wiksell.
Kushida, H., & Okada, M. (2003). A proof-theoretic study of the correspondence of classical logic and modal logic. *Journal of Symbolic Logic, 68*(4), 1403–1414. https://doi.org/10.2178/jsl/1067620195
Lehrer, K. (2018). *Theory of knowledge* (2nd ed.). New York: Routledge.
Lehrer, K., & Paxson, T., Jr. (1969). Knowledge: Undefeated justified true belief. *The Journal of Philosophy, 66*(8), 225–237.
Lenzen, W. (1979). Epistemologische betrachtungen zu [S4, S5]. *Erkenntnis, 14*(1), 33–56.
Lewis, D. (1973). *Counterfactuals*. Cambridge, MA: Harvard University Press.
Maffezioli, P., Naibo, A., & Negri, S. (2013). The Church-Fitch knowability paradox in the light of structural proof theory. *Synthese, 190*(14), 2677–2716.
Massacci, F. (2000). Single step tableaux for modal logics. *Journal of Automated Reasoning, 24*(3), 319–364.
Mints, G. (1997). Indexed systems of sequents and cut-elimination. *Journal of Philosophical Logic, 26*(6), 671–696.
Montague, R. (1970). Universal grammar. *Theoria, 36*(3), 373–398. https://doi.org/10.1111/j.1755-2567.1970.tb00434.x
Negri, S. (2005). Proof analysis in modal logic. *Journal of Philosophical Logic, 34*, 507–544. https://doi.org/10.1007/s10992-005-2267-3
Negri, S. (2017a). Non-normal modal logics: A challenge to proof theory. In P. Arazim & T. Lávička (Eds.), *The Logica Yearbook 2016* (pp. 125–140). College Publications.
Negri, S. (2017b). Proof theory for non-normal modal logics: The neighbourhood formalism and basic results. *The IfCoLog Journal of Logics and their Applications*, *4*(4), 1241–1286. Correction: *4*(4), xi–xiv.
Negri, S., & Olivetti, N. (2015). A sequent calculus for preferential conditional logic based on neighbourhood semantics. In H. De Nivelle (Ed.), *Lecture notes in computer science* (pp. 115–134). Cham: Springer International Publishing.
Negri, S., & von Plato, J. (2011). *Proof analysis: A contribution to Hilbert's last problem*. Cambridge: Cambridge University Press. https://doi.org/10.1017/CBO9781139003513
Nerode, A. (1991). Some lectures on modal logic. In F. L. Bauer (Ed.), *Logic, algebra, and computation* (Vol. F79, pp. 281–334). Springer.
Nomura, S., Ono, H., & Sano, K. (2020). A cut-free labelled sequent calculus for dynamic epistemic logic. *Journal of Logic and Computation, 30*(1), 321–348. https://doi.org/10.1093/logcom/exaa014
Orlowska, E., & Golińska Pilarek, J. (2011). *Dual tableaux: Foundations, methodology, case studies* (Vol. 33). Dordrecht: Springer.
Pacuit, E. (2013). Dynamic epistemic logic I: Modeling knowledge and belief. *Philosophy Compass, 8*(9), 798–814. https://onlinelibrary.wiley.com/doi/abs/10.1111/phc3.12059

Pacuit, E. (2017). *Neighborhood semantics for modal logic*. Cham: Springer International Publishing.
Priest, G. (2008). *An introduction to non-classical logic: From if to is*. Cambridge: Cambridge University Press.
Rott, H. (2004). Stability, strength and sensitivity: Converting belief into knowledge. In E. Brendel & C. Jäger (Eds.), *Contextualisms in epistemology* (pp. 327–351). Dordrecht, Netherlands: Springer.
Russell, B. (1912). *The problems of philosophy*. New York: Henry Holt and Company.
Russo, A. (1995). Modal labelled deductive systems. *Department of Computing, Imperial College, London, Technical Report*, 9571–9599.
Scott, D. (1970). Advice on modal logic. In K. Lambert (Ed.), *Philosophical problems in logic: Some recent developments* (pp. 143–173). Dordrecht: Springer Netherlands. https://doi.org/10.1007/978-94-010-3272-8_7
Simpson, A. (1994). *The proof theory and semantics of intuitionistic modal logic*. University of Edinburgh, School of Informatics.
Sosa, E. (1964). The analysis of 'knowledge that p'. *Analysis, 25*(1), 1–8.
Sosis, C. (2017). *What is it like to be a philosopher?—An interview with Timothy Williamson*. Retrieved September 03, 2021, from http://www.whatisitliketobeaphilosopher.com/timothy-williamson/
Stalnaker, R. (1996). Knowledge, belief and counterfactual reasoning in games. *Economics and Philosophy, 12*(2), 133–163.
Stalnaker, R. (1998). Belief revision in games: Forward and backward induction. *Mathematical Social Sciences, 36*(1), 31–56. http://www.sciencedirect.com/science/article/pii/S0165489698000079
Stalnaker, R. (2006). On logics of knowledge and belief. *Philosophical Studies, 128*(1), 169–199.
Swain, M. (1974). Epistemic defeasibility. *American Philosophical Quarterly, 11*(1), 15–25.
van der Hoek, W. (1993). Systems for knowledge and belief. *Journal of Logic and Computation, 3*(2), 173–173.
Van Ditmarsch, H., van Der Hoek, W., & Kooi, B. (2008). *Dynamic epistemic logic* (Vol. 337). Springer Science & Business Media.
Viganó, L. (2000). *Labelled non-classical logic*. Boston/Dordrecht/London: Kluwer.
Williamson, T. (2001). Some philosophical aspects of reasoning about knowledge. In *Proceedings of the 8th conference on Theoretical Aspects of Rationality and Knowledge*. (p. 97).
Williamson, T. (2002). *Knowledge and its limits*. Oxford: Oxford University Press on Demand.
Williamson, T. (2018). *Doing philosophy: From common curiosity to logical reasoning*. Oxford: Oxford University Press.

# Chapter 13
# Mission Impossible

**Graham Priest**

**Abstract** Saul Kripke's work on the semantics of non-normal modal logics introduced the idea of non-normal worlds, worlds where certain connectives behave differently from the way in which they behave in the worlds of normal modal logics. Such worlds may be thought of as impossible worlds, though Kripke did not, himself, talk of them in this way. Since Kripke's invention, the notion of an impossible world has undergone much fruitful development and application. Impossible worlds may be of different kinds—or maybe different degrees of impossibility; and these worlds have found application in many areas where hyperintensionality appears to play a significant role: intentional mental states, counterfactuals, meaning, property theory, to name but a few areas. But what, exactly, is an impossible world? How is it best to characterise the notion? To date, the notion is used more by example than by definition. In this paper I will investigate the question, and propose a general characterisation, suitable for all standard purposes and tastes.

**Keywords** Closure under logical consequence · Extrinsic logics · Formal modal semantics · Impossible worlds · Non-normal worlds

*Dedication*: It gives me great pleasure to dedicate this essay to Saul. Without his work, contemporary studies of non-classical logic would be unthinkable. It has been an inspiration for all who work in logic.

G. Priest (✉)
The CUNY Graduate Center, New York, USA
e-mail: priest.graham@gmail.com

The University of Melbourne, Melbourne, Australia

The Ruhr University of Bochum, Bochum, Germany

© The Author(s), under exclusive license to Springer Nature Switzerland AG 2024

Y. Weiss and R. Birman (eds.), *Saul Kripke on Modal Logic*, Outstanding Contributions to Logic 30, https://doi.org/10.1007/978-3-031-57635-5_13

## 13.1 Introduction

Saul Kripke's work on the semantics of non-normal modal logics introduced the idea of non-normal worlds, worlds where logically impossible things may hold. Such worlds can naturally be thought of as impossible worlds. Since Kripke's invention in Kripke (1965), the notion of an impossible world has undergone much fruitful development and application. Impossible worlds may be of different kinds—or maybe different degrees of impossibility (for example, they may be closed under conditions of different degrees of strength, or none at all); and these worlds have found application in many areas where hyperintensionality appears to play a significant role: intentional mental states, counterfactuals, meaning, property theory, to name but a few areas.

But what, exactly, is an impossible world? How is it best to characterise the notion? To date, the notion is used more by example than by definition. In this paper I will investigate the question and propose a general characterisation, suitable for all standard purposes and tastes. It will be:

- independent of all applications
- neutral on the ontological question of what worlds, as such, are
- neutral on the question of what inferences are valid
- neutral on the question of what it is that makes them so.

In particular, it can be deployed whatever one takes the correct logic to be.

The guiding thought of what follows is this[1]:

- a possible world is one that is closed under logical consequence, and an impossible world is one that is not.

However, this thought needs to be articulated in different ways, depending on whether one is dealing with the worlds of a formal logical semantics or worlds in reality themselves—whatever this means. In the first case, we have a logical consequence relation to hand, given by the formal semantics. In the second, we do not: we have to extract it from the worlds themselves, as it were.

In their (2018) and at greater length in their (2019, Sect. 1.4), Berto and Jago suggest four possible definitions of an impossible world: (1) a world where something impossible happens; (2) a world where the laws of logic fail; (3) a world where the logic is non-classical; (4) a world where contradictions hold. The account I shall provide is a version of (2). I reject (1) for reasons that we will come to in Sect. 13.5.1; (3) presupposes the truth of classical logic, whereas I seek an account which does not prejudge that issue. A similar comment applies to (4), even if it were not already

[1] This definition is suggested, in effect, in Priest (1997a, p. 482).

problematic because a failure of $A \vee \neg A$ is just as impossible in classical logic as a holding of $A \wedge \neg A$.[2]

The paper has the following structure. In the next section I provide a little historical background on worlds, possible and impossible. Given what I have to say, it behoves me to explain what I take a logic to be. This is spelled out in the next section. In Sect. 13.4, I show how the guiding thought of the paper is to be articulated in the case of the worlds of a formal semantics. Then in Sect. 13.5, I turn to the much harder question of how it is to be articulated in the case of worlds themselves. In Sect. 13.6, I note a technical iteration of the material in Sect. 13.5. Then in the final section, Sect. 13.7, I tie up a number of the loose ends which have appeared in the discussion.

## 13.2 Background: "Kripke Semantics"

### *13.2.1 Possible Worlds*

First, then, some historical background.

Kripke's work on the semantics of modal logic was a game-changer. Modal logic was studied in the West in both Ancient and Medieval logic, but the topic had largely fallen into oblivion by the start of the 20th Century. It was rescued from this state in the early part of the century by the work of C. I. Lewis, who proposed—famously—his five axiomatic systems of modal logic, *S*1–*S*5. Modal and other intentional notions came under heavy attack by Quine and others around the middle of the century, and I think it fair to say that by the 1960s they were under a deep cloud.

At about this time, Carnap, Prior, and others, were working with ideas that could provide a semantics for modal logic[3]; but it was Kripke's papers Kripke (1959) and Kripke (1963), deploying explicitly the notion of a possible world, which really established such semantics as logically kosher. It is no accident that this genre of semantics is now termed 'Kripke Semantics'. The Quinean skepticism did not go away immediately, and certainly the metaphysics of possible worlds caused many brows to furrow. However, the world semantics found so many applications, not only concerning modal notions, but also other intensional notions—such as belief and knowledge, counterfactuals, meaning, property theory, and a number of other

[2] Zalta (1997) gives a different definition of worlds, possible and impossible; but it, too, is not logic-neutral. For him, all worlds (possible and impossible) have to be complete; but in paracomplete logics even possible worlds do not have to be complete. And for him, inconsistent worlds are impossible; but in paraconsistent logics, inconsistent worlds can be possible.

[3] There was also a slightly earlier algebraic semantics for some modal logics in the work of McKinsey and Tarski (1944, 1946); but I think it fair to say that this was seen as just a bit of pure mathematics.

areas—that in due course Quinean scruples quietly disappeared. The techniques were just too powerful and versatile not to be used—though philosophers could still, and do still, argue about what to make of possible worlds metaphysically.[4]

### *13.2.2 Impossible Worlds*

The Lewis systems $S4$ and $S5$ are now termed *normal modal logics*. $S4$ and $S5$ are but two of a large family of such logics uncovered by Kripke semantics, and not the most fundamental. However, the systems $S1$, $S2$, and $S3$ are not normal systems. Something that distinguishes them syntactically is that the Rule of Necessitation, $\vdash A \Rightarrow \vdash \Box A$, fails. In yet another of his papers, Kripke (1965) provided a semantics for these.[5] To provide these semantics, Kripke introduced the idea of a different kind of world, which he called *non-normal*. Hence, these systems are now usually known as non-normal modal logics. At non-normal worlds, the truth conditions of formulas of the form $\Box A$ and $\Diamond A$ were not given in terms of a binary accessibility relation, as for normal worlds. Rather, if $w$ is a non-normal world, then every formula of the form $\Box A$ is false, and every formula of the form $\Diamond A$ is true at $w$.[6] This means that logical truths, such as $\Box(A \vee \neg A)$, may fail at such worlds; and logical falsities, such as $\Diamond(A \wedge \neg A)$, may hold. This makes it natural to think of such worlds as logically impossible worlds—though Kripke did not refer to them as such.

Kripke's impossible worlds were of a very simple kind, but many more applications of the notion of an impossible world were soon on the market. In particular, they play a central role in the semantics of relevant logic[7]—though again, they were not conceptualised as impossible worlds till some time after their introduction.[8] In these semantics, if $A$ is *any* logical truth, it may fail at some impossible worlds; and if $A$ is *any* logical falsity, it may hold at some impossible worlds. The impossible worlds of relevant logic are still fairly orderly, however. In particular, they are closed under *modus ponens* for logical truths of the form $A \to B$. A use for much more anarchic impossible worlds came a little later in the semantics of intentional notions. Here, virtually *any* combination of formulas may hold or may fail.[9]

---

[4] References: On the history of modal logic, see Knuuttila (2012). On the genesis of contemporary modal logic, see Ballarin (2022). On the Quinean attack on modal logic, see Barcan Marcus (1995). On possible worlds, see Menzel (2016).

[5] This is a slight exaggeration. The semantics were for $S2$ and $S3$. $S1$ proved to be more recalcitrant, and semantics were not found till later. See Cresswell (1972).

[6] See Priest (2008, Chap. 4).

[7] See Routley et al. (1982).

[8] As far as I know, the first place where this terminology appears in print in this context is Priest (1992).

[9] See, e.g., Priest (2005), Bjerring (2013), Berto (2017).

Indeed, there are now many applications of semantics with impossible worlds.[10] And I think it fair to say that if worlds are to be used for the analysis of many notions, especially hyperintensional notions, some of those worlds must be impossible worlds.[11]

Of course, the addition of impossible worlds of various kinds to the menagerie of worlds has complicated debates about the metaphysics of worlds. But this is not my concern here.[12] And of course, people have objected to impossible worlds. This, also, is not my concern here.[13] My concern is much more basic than these matters, and is this: What, exactly, is an impossible world? What is it about a world that makes it impossible? And let me emphasize right at the start, that I am talking about logical impossibility, not some other kind of impossibility.

## 13.3 What Is a Logic?

Let us now turn to matters of philosophical substance, starting with the question of what, for the present purposes, we may take a logic to be—and in this essay, I will mean by 'logic' *deductive logic* (since the debate about possible and impossible worlds arises in the context of such logics).

We may start by fixing a language, $\mathfrak{L}$. For present, we may assume that this is some standard propositional language. We may now identify a logic with a consequence relation on $\mathfrak{L}$, $\vdash$. That is, it is a relation, $\Pi \vdash \Gamma$, where $\Pi$ and $\Gamma$ are sets of formulas

[10] See Priest (1997a), the other essays in Priest (1997b), Berto and Jago (2018), and Parts II and III of Berto and Jago (2019). See also Berto (2017), Berto et al. (2018), Bjerring (2013, 2014), Brogard and Salerno (2013), Jago (2013), Krakaur (2013), Nolan (2013), Sillari (2008), Stuart, McLoone and Grützner (2023), Weber and Omori (2019), Yagisawa (2010).

[11] See, e.g., Nolan (2013, 2014).

[12] The matter is discussed in Berto and Jago (2018), and Part I of Berto and Jago (2019). For what it's worth, I take all worlds other than the actual world to be non-existent objects. See Priest (2005, Sect. 7.3).

[13] The matter is also discussed in Berto and Jago (2018). Again for what it is worth, Berto and Jago note six objections that have been made. I have nothing new to say about the first five of these. But I do about the sixth objection (which is discussed at greater length in Berto and Jago (2019, esp. Sect. 8.5)). This is to the effect that at impossible worlds the truth value(s) of a formula do not depend on the truth value(s) of its sub-formulas (maybe at other worlds). In that sense, the semantics is not compositional. And many have felt that a semantics should be compositional to account for learnability. I take this to be a mistaken thought. What learnability requires is that the semantics be recursive, or at least recursively enumerable, which the semantics with impossible worlds are, or may be. It is true that in the most general case the specification of the value of each formula at an impossible world must be fixed independently; but the same is true, of course, of propositional parameters. It might be suggested that in any learnable language there should be only a finite number of propositional parameters; if so, in the modal case, one might just restrict oneself to a language with formulas of some finite complexity (which is required, in any case, if the semantics are to be humanly learnable). Alternatively, one may simply restrict oneself to specifications of truth values of formulas at impossible worlds which are recursively enumerable.

of $\mathfrak{L}$.[14] Intuitively, one may think of this as meaning something like: if all of the members of $\Pi$ hold, then some of the members of $\Gamma$ hold. In the most general case, $\Pi$ and $\Gamma$ may be infinite, but provided we are dealing with a compact consequence relation, the infinite case may be defined from the finite case in the familiar way.[15] Since non-compact logics have played little role in modal logics, we will restrict our attention to finite $\Pi$ and $\Gamma$ here. I will make a brief comment on the more general case in Sect. 13.7.2.

It is standard to put further constraints on $\vdash$. Thus, the "Tarski conditions" are often imposed. $\vdash$ is Reflexive:

- $\Sigma \cup \{A\} \vdash \{A\} \cup \Delta$

Monotonic:

- if $\Sigma \vdash \Delta$, $\Theta \supseteq \Sigma$, and $\Delta \subseteq \Xi$, then $\Theta \vdash \Xi$

and Transitive:

- if $\Sigma \vdash \Delta \cup \{A\}$ and $\{A\} \cup \Theta \vdash \Xi$ then $\Sigma \cup \Theta \vdash \Delta \cup \Xi$

Standard definitions of validity using world semantics deliver all these conditions. So I will assume them here. Again, I will make a brief comment on relaxing these conditions in Sect. 13.7.1.

Another constraint often imposed on a consequence relation is closure under substitution. That is, let $s$ be any function from propositional parameters to the formulas of $\mathfrak{L}$. If $A$ is any formula of the language, let $A_s$ be $A$ with every propositional parameter, $p$, replaced by $s(p)$. Let $\Sigma_s = \{A_s : A \in \Sigma\}$. Then closure under substitution is the condition:

- if $\Pi \vdash \Gamma$ then $\Pi_s \vdash \Gamma_s$

This closure condition will surface briefly in what follows, but I will not pack it into the definition of a consequence relation.

Finally, if the language has conjunction and disjunction operators which work in the usual fashion, one may reduce the finite non-empty case to the singleton case as follows: $\bigwedge\{\pi \in \Pi\} \vdash \bigvee\{\gamma \in \Gamma\}$.[16] However, we will not assume this here, since such conditions can fail in logics with world-semantics. Thus, if supervaluation is employed, one may have $A \vee B \vdash A \vee B$, but not $A \vee B \vdash A, B$; and if subvaluation is employed, one may have $A \wedge B \vdash A \wedge B$, but not $A, B \vdash A \wedge B$.[17]

---

[14] Those who envisage a world-semantics for inferences whose premises/conclusions are multisets (or some other type of data-structure), are free to make appropriate changes. As is standard, I will omit the set brackets when the sense is clear without them.

[15] $\Pi \vdash \Gamma$ iff for some finite subsets, $\Pi' \subseteq \Pi$ and $\Gamma' \subseteq \Gamma$, $\Pi' \vdash \Gamma'$.

[16] And one can reduce the empty case to the singleton case if the language contains the constants $\bot$ and $\top$, the first entailing everything, and the second being entailed by everything.

[17] See Priest (2008, Sect. 7.10.).

## 13.4 Formal Semantics

We may now turn to our question: What is an impossible world? We need to distinguish here between the worlds of a formal semantics, and whatever it is that these are supposed to represent in reality. The answer concerning the first is a lot easier than the answer concerning the second. So let us take this first.

We suppose that we have a world-semantics for some logic, $\vdash$, such that validity is defined as truth—however that is cashed out—preservation over a subset (possibly improper) of worlds. Worlds are simply mathematical entities of a certain kind—perhaps numbers or sets. Which are the possible ones? The obvious answer is that they are the ones over which validity is defined. But how might these be characterised intrinsically? We can answer this by considering the relation $w \vDash A$, expressing the claim that $A$ is true at $w$.

Given this, we can define a possible world (with respect to $\vdash$) as one which is closed under $\vdash$. That is, if $w$ is one of the worlds at issue, $w$ is possible iff:

- if $\Pi \vdash \Gamma$ and $w \vDash A$ for all $A \in \Pi$, then $w \vDash B$ for some $B \in \Gamma$

A world is impossible if it is not possible.

Note that the definition makes no assumptions about what $\vdash$ is, and in that sense is logic-neutral. However, as one would expect, given that we are dealing with a pre-given semantics, the possible and impossible worlds at issue are not, since these will depend on whatever logic is at issue. The collection of worlds itself will also be determined by the formal semantics.

This definition of 'impossible world' is entirely unproblematic in most cases; however, it can give a few surprising results. For a start, in some semantics (for a logic, $\vdash$), a world that is officially classed as impossible (non-normal), may actually be possible according to this definition. For example, if what is true at such a world is entirely arbitrary, it can be closed under $\vdash$ by accident, as it were.[18] In that case, one can just move the world into the official class of possible worlds without changing the consequence relation.

Perhaps more interestingly, there are logics given by the Kripke semantics for $S2$ and $S3$, but where validity is defined over all worlds (not just non-normal ones).[19] Contrary to what one might have expected, as defined, all the worlds are possible. There are just two classes of these, normal and non-normal. (Modal formulas simply have disjunctive truth conditions.[20]) $S2$ and $S3$ notwithstanding, in these logics, since validity is preserved at all worlds, there are none at which logically aberrant things happen.

One might think that the opposite can also happen: that some officially possible worlds are impossible. Thus, in some semantics there is a trivial world, where everything is true, and/or (if there are no logical truths) an empty world where nothing is

[18] See Priest (2005, 2nd edn, Chap. 9).

[19] These are the Lemmon "$E$ Systems". See Priest (2008, 4.4.5).

[20] See, further, Priest (2005, 2nd edn, 13.3.2).

true.[21] These are possible according to this definition, though one might think them impossible. Actually, I think that in logics delivered by the semantics, these worlds *are* possible: nothing happens which is ruled out by the logic. However, if one does not like this, one could always treat these worlds as a special cases, as impossible by definition.

Finally, I note also that although an impossible world is not closed under $\vdash$, it may be closed under some weaker condition. Thus, in relevant logics, all worlds (possible and impossible) are closed under *modus ponens* for logical truths of the form $A \to B$.[22] However, in a general semantics for intentional operators, if $w$ is an impossible world, the set $\{A : w \vDash A\}$ may be quite arbitrary.[23] The amount of anarchy of an impossible world may therefore come by degrees.

## 13.5 Worlds Themselves

We now turn to the much tricker problem of defining what an impossible world is when we are dealing with worlds themselves. That is, we are not dealing simply with the mathematical machinery of some formal semantics, but the worlds *in re*—whatever that means.

We have an actual world, @, and a bunch of other worlds. $\mathfrak{L}$ is a language adequate for describing what holds at words, with—perhaps—the exception of the facts about logical consequence. However, now that we have no model theory, and so no notion of an interpretation to deploy, $\mathfrak{L}$ has to be taken as an interpreted language.

To say which worlds are possible/impossible, we will apply again the guiding thought of the essay: that a possible world is one which is closed under logical consequence. When dealing with the worlds of a formal semantics, we have a consequence relation ready to hand; but now this luxury is no longer available to us. And let me emphasise again that we are talking about *logical* impossibility—impossibility according to the laws of logic. There are, of course, other kinds. Indeed, we will meet one in a moment.[24]

I think that the beginning of an answer is to be found by analogy with physical impossibility. A physically impossible world is one where the laws of nature are different from those of the actual world. In such a world, some physical event or state of affairs which cannot happen at our world, can happen; or which can happen at our world, cannot happen. By analogy, a logically impossible world is one where the

[21] See, for example, Weber and Omori (2019).

[22] See Priest (2008, Chap. 10).

[23] See Priest (2005, Chap. 1).

[24] Some hold that there is a distinctive notion of metaphysical impossibility—that for example, it is metaphysically impossible that Hesperus is Phosphorus, though this is not a logical impossibility. Actually, I am skeptical of this. (See Priest (2021).) However, in any case, that notion of necessity is not what is at issue here.

laws of logic are different from the laws of logic at the actual world.[25] The definition is neutral as to what that logic—"the correct logic"—is. The definite description does, however, presume that a world has a unique logic. So be it for the moment. I will come back to the matter in Sect. 13.7.3.

Of course, how one understands what it means to say that such and such are the laws of nature of a world is a somewhat contentious question.[26] However, it is not necessary to plunge into those murky depths here. Clearly, we *do* have to engage with the more unfamiliar question of what it means for such and such to be the logical laws of a world. To the extent that we need to answer the question, we can do this in terms of the relation $w \vDash A$, for the worlds, $w$, in question. However, now that we are not dealing with a formal semantics, we cannot presuppose a definition provided by such. On how, in fact, the notion is to be analysed we need take no position, and so is another matter on which the account that follows is neutral.

### *13.5.1 A World's Extrinsic Logics*

Since logic is at least something that is truth preserving, for a consequence relation, $\vdash$, to be the logic of a world, what is true at the world should be closed under $\vdash$. That is, for $\vdash$ to be the logic of $w$ we require that:

- if $\Pi \vdash \Gamma$ and $w \vDash A$ for all $A \in \Pi$, then $w \vDash B$ for some $B \in \Gamma$

If this condition is satisfied, let us then say that $\vdash$ is an *extrinsic logic* of $w$.

Clearly, an extrinsic logic of a world is not unique, since if $\vdash' \subseteq \vdash$ and $\vdash$ is an extrinsic logic of $w$, so is $\vdash'$. The obvious thought then is to define the logic of the world as one of the extreme cases: either the weakest truth preserving relation, or the strongest. But the weakest is the null relation, which relates nothing to anything; and the strongest is the one that is materially truth preserving. These have no claim to be a logic in any substantial sense.

Even if we set these extreme cases aside, there are inferences that are necessarily truth preserving but are not usually taken to be valid. Thus, the inference '$a$ is red; so $a$ is coloured' is truth preserving at possible worlds. It is not normally taken to be a deductively valid inference. One may rule out this sort of situation by insisting that in an extrinsic logic, not only are the inferences truth preserving, but their substitution instances are truth preserving as well. This also fails, however: such inferences may still be truth-preserving "by accident". Compare the case with that of a law of nature. Suppose that the laws of motion are those of Special Relativity, so that nothing accelerates through the speed of light. Now consider a world where the laws of motion are Newtonian, but which doesn't last very long, and where everything moves pretty slowly, so that nothing has time to reach the velocity of light. At such a world it is also true that nothing accelerates through the speed of light. However,

[25] The definition is mooted in Priest (1992, Sect. 13.2).

[26] On how one might understand the notion of a law of nature, see Carroll (2020).

this is simply a contingent feature of the world, and it does not make the laws of motion of the world those of Special Relativity. That is, the universal generalisation is contingent and not lawlike.

Similar things may also happen in the case of logic.[27] If $\vdash$ is the logic of @, it should be truth preserving not just at @, but at all possible worlds—whatever they are. Now, suppose that there are possible worlds where contradictions are true, but that @ is a consistent world. Then any inference of the form: '$p \wedge \neg p$; so $q$' is truth preserving at @, but not at all possible worlds. A similar sort of situation can arise in classical logic, as was pointed out by Etchemendy.[28] Suppose that it is a contingent fact that the actual world has only one object, but that some possible worlds are larger. Then '$p$; so $\exists x \forall y\, x = y$' is truth preserving at @, but not at all possible worlds. A third example of the same phenomenon arises if one has a world semantics with an actuality operator, $\hat{@}$, in the language, such that for any (possible) world, $w$, $w \vDash \hat{@}A$ iff $@ \vDash A$.[29] For then '$A$; so $\hat{@}A$' is truth preserving at the actual world, but not at all possible worlds.

### *13.5.2 The Intrinsic Logic of a World*

In general, then, we have a collection of consequence relations which, for all we have seen so far, might be the logic of a world. Given that we are taking each world to have a unique logic, more needs to be said. A way to do so is by making reference to what each world "thinks" its own logic to be. In other words, one needs the language to contain the means to express the laws of validity.

In the case of physical laws, we can express the fact that a universal generalisation is a law of nature (lawlike), and not a contingent generalisation, by using modal operators. Thus, in a world where the laws of motion are Newtonian, an object *can* (physically) accelerate through the speed of light. In a world where the laws are those of Special Relativity, it *cannot*.

A similar trick will not work in the case of laws of logic, since modal distinctions will not suffice for the task. Thus, in many logics (e.g., $LP$, $FDE$), models may have a trivial world, where everything holds. So if the modal accessibility relation accesses this, $\vDash \Diamond A$ is true for all $A$, even though these are different logics.

The trouble is that logical consequence is essentially relational, and cannot be reduced to something non-relational. In particular, even in just the one-premise one-conclusion case, one cannot characterise $A \vdash B$ by something like $\neg\Diamond(A \wedge \neg B)$. In relevant logics, for example, $p \wedge (p \to q) \vdash q$, but there may be (consistent)

[27] So there is a difference between a world where the laws of logic are different (from the actual laws), and a world where the logically impossible happens. I take the former to be the criterion of an impossible world, for the reasons just given. If I understand his account correctly, Tanaka (2018) endorses the latter.

[28] Etchemendy (1990, pp. 111 ff).

[29] See Crossley and Humberstone (1977).

modally accessible worlds, not closed under *modus ponens*, where $p \wedge (p \rightarrow q)$ and $\neg q$ both hold.

Hence, to express logical consequence, we need a notion that is essentially relational. To this end, let us extend our language with a new connective, $\Rightarrow$, such that $A_1, ..., A_m \Rightarrow B_1, ..., B_n$ is a formula, where the left and right hand sides are finite enumerations (possibly empty) of formulas of $\mathfrak{L}$.[30] Call the extended language $\mathfrak{L}^{\Rightarrow}$. If you want to read $\Rightarrow$ in English, it's something like: given that $A_1$, ..., and that $A_n$ then it follows that that $B_1$, ..., or that $B_m$.[31] Notice that the connective is multigrade, in that it may have a variable number of arguments both left and right of the $\Rightarrow$.[32] (As we shall note in due course, if we assume that the language contains vocabulary for talking about sets, we might take what we need to be expressed by a relation between two sets, not a connective between formulas. However, I make no such assumption here.)

$\Rightarrow$-statements are true or otherwise at a world, just as are the statements of $\mathfrak{L}$. As to what makes them so, we may take no stand here. But let us say that the relation between two sets of formulas, $\vdash_w$, defined by:

- $A_1, ..., A_m \vdash_w B_1, ..., B_n$ iff $w \vDash A_1, ..., A_m \Rightarrow B_1, ..., B_n$

is the *intrinsic* logic of the world. *Warning*: there is no guarantee that $\vdash_w$ *is* a logic as we have defined it. Indeed, there is no reason that it must even be truth preserving at $w$. Let us call a world where it is, *coherent*. That is, a world, $w$, is coherent if $\vdash_w$ is an extrinsic logic of $w$.

Since truth preservation is a necessary constraint on the (actually) true theory of validity, the actual world, @, is coherent. (Though, of course, which $\Rightarrow$-sentences *are* true is a contentious matter of the correct logic, on which we also take no stand here.)

And now, at last, we can provide the required definitions. $w$ is a possible world if it is coherent and has the same intrinsic logic as @. (So, in particular, @ is a possible world.) It is impossible otherwise. Thus, the actual world and its intrinsic logic determine which of the other words are possible/impossible.[33]

Let me finish this section with three observations concerning the definition.

---

[30] As already noted, I make no assumption about the behaviour of conjunction and disjunction if they are in the language. So this cannot be reduced to the single premise/conclusion case.

[31] I note that even if we assume that the commas can be coded as conjunctions and disjunctions, in no standard logic with a world semantics is there a connective, $\rightarrow$, such that we are guaranteed that for a possible world, $w$, $w \vDash (A_1 \wedge ... \wedge A_n) \rightarrow (B_1 \vee ... \vee B_m)$ iff $A_1 \wedge ... \wedge A_n \vDash B \vee ... \vee B_m$. If the logic has the "deduction theorem" for $\rightarrow$ we will have the right to left direction, but even then, nothing guarantees the left to right direction.

[32] One usually meets the characteristic of being multigrade in connection with predicates, not connectives. See Grandy (1976) and Taylor and Hazen (1992). However, multigrade connectives are known. See Humberstone (2011, pp. 630, 783) for discussions and references. See also Humberstone (2021). Taylor and Hazen operate with predicates of a fixed arity, each of whose places can be filled by an arbitrary number of arguments. This is essentially how the connective $\Rightarrow$ functions. One might call such predicates/connective 'polyunsaturated'.

[33] Clearly, any other coherent world could play the same role in this definition. So we could define a notion of possible-with-respect-to $w$, for each coherent world, $w$. Hence there will be sets of

First, assuming that $\Box A$ is true iff $A$ is true at all possible worlds. Then, as one would expect, if $@ \vDash A_1, ..., A_m \Rightarrow B_1, ..., B_n$, $@ \vDash \Box(A_1, ..., A_m \Rightarrow B_1, ..., B_n)$.

Secondly, impossible worlds may be of different kinds. There are coherent ones, where truth is closed under the intrinsic logic of the world; it is just that that logic is not the same as that of @. But there are also incoherent worlds, where true $\Rightarrow$-statements do not preserve truth. One might already have expected there to be such worlds. After all, at an impossible world, anything might happen![34]

Finally, it is quite possible that two coherent worlds have all the same extrinsic logics, but different intrinsic logics. Analogously for physical necessity, two worlds might have all the same true universal generalisations, but which ones are lawlike may be different in different worlds.[35] Thus, two worlds may agree on what holds in the language $\mathfrak{L}$, whilst differing in what holds in the language $\mathfrak{L}^{\Rightarrow}$. In the same way, the description of a finite mathematical structure might be the same at both a world whose logic is classical and a world whose logic is intuitionist. However, in one world $\Rightarrow A \vee \neg A$ may hold, but not at the other.

## 13.6 Higher Orders

The preceding considerations are sufficient to answer the central question of this essay. But a question will naturally occur to any logician at this point.

In Sect. 13.5, the language $\mathfrak{L}^{\Rightarrow}$ expresses a consequence relation between formulas in the language $\mathfrak{L}$. But we have extended $\mathfrak{L}$ to do this. So there is now the matter of the consequence relation between formulas of the extended language.[36]

To get a sense of what is at issue here, note that two of the Tarski closure conditions (Monotonicity and Transitivity) are themselves naturally framed as such inferences. Thus, Transitivity—or Cut, as it is more usually called in this context—can be expressed by the inference:

$$\frac{\Sigma \Rightarrow \Delta \cup \{A\} \quad \{A\} \cup \Theta \Rightarrow \Xi}{\Sigma \cup \Theta \Rightarrow \Delta \cup \Xi}$$

classical possible worlds, intuitionistic possible worlds, etc. However, real possibility is defined with respect to the actual world. And for better or for worse, there is only one of these, even if actuality is indexical, and every world takes itself to be actual.

[34] Of course, incoherence may come by degrees. Suppose that the world structure is given by that of some standard non-normal modal logic. Then the impossible worlds are closed under classical propositional logic, but not modal validities (e.g., $A \Rightarrow \Box(B \vee \neg B)$.).

[35] As even "neo-humeans" about physical necessity like David Lewis agree. (See Carroll (2020, Sect. 13.2).)

[36] In truth, there is absolutely nothing in the construction which assumes that $\Rightarrow$ is not a part of $\mathfrak{L}$. However, I have not assumed this, since the connective $\Rightarrow$ is not of a kind that is familiar in logic. But see Sect. 13.7.2.

The facts about $\Rightarrow$ do not determine the facts about these meta-inferences. To see this, just note that the Cut Theorem for classical logic shows us that the same inferences are valid both with and without Cut.

Bearing these things in mind, we may now extend the construction used for $\mathfrak{L}$ to sentences of the language $\mathfrak{L}^{\Rightarrow}$. We can, in fact, proceed exactly as before at this higher level. Say that a consequence relation at this level, $\vdash$, is an extrinsic logic of world $w$ if:

- if $\Pi \vdash \Gamma$ and $w \vDash A$ for all $A \in \Pi$, then $w \vDash B$ for some $B \in \Gamma$

where all of the formulas in $\Pi$ and $\Gamma$ contain $\Rightarrow$.[37] We now have to extend the language to express the intrinsic logic at one level up.

To keep track of things in what follows, let us write what we have hitherto written as $\Rightarrow$, as $\Rightarrow^0$, and call what has gone before, level 0. We now add a new connective to the language, $\Rightarrow^1$, such that $A_1, ..., A_m \Rightarrow^1 B_1, ..., B_n$ is a formula, where all of the $A$s and $B$s contain $\Rightarrow^0$. An instance of Cut might then be expressed by[38]:

$$(\Sigma \Rightarrow^0 \Delta \cup \{A\}, \{A\} \cup \Theta \Rightarrow^0 \Xi) \Rightarrow^1 (\Sigma \cup \Theta \Rightarrow^0 \Delta \cup \Xi)$$

Call this language $\mathfrak{L}^{\Rightarrow^1}$. Let us say that the relation $\vdash^1_w$ such that:

- $A_1, ..., A_m \vdash^1_w B_1, ..., B_n$ iff $w \vDash A_1, ..., A_m \Rightarrow^1 B_1, ..., B_n$

is the intrinsic logic of $w$ at level 1. A world is coherent at this level if the intrinsic logic is an extrinsic logic. @ is coherent at this level for the same reason that it was coherent at level 0. A world is now possible if it is coherent at levels 0 and 1, and its intrinsic logics are the same as those of @ at both levels.[39]

Of course, we now face the same situation with respect to inferences between sentences of $\mathfrak{L}^{\Rightarrow^1}$, and so on for higher levels. As I noted, the facts about $\Rightarrow^0$ do not uniquely determine the facts about $\Rightarrow^1$. Exactly the same is true at higher levels.[40] We may then just iterate our construction. The general procedure is this.

Let $\mathfrak{L}^0$ be $\mathfrak{L}$, and let $\mathfrak{L}^{i+1}$ be $\mathfrak{L}^i$ augmented with the connective $\Rightarrow^i$. For all $i > 1$, say that a consequence relation, $\vdash$, is an extrinsic consequence relation at level $i$ for world $w$ if:

- if $\Pi \vdash \Gamma$ and $w \vDash A$ for all $A \in \Pi$, then $w \vDash B$ for some $B \in \Gamma$

---

[37] Note that this construction does not allow for (the expression of) inferences of "cross type". With a bit of fiddling, we could do so by allowing for a more liberal syntax. Similar comments apply to what follows.

[38] I insert brackets for ease of readability. The set terms are to be thought of as the appropriate enumerations.

[39] In fact, we could take the intrinsic logic of a world (*sans* level) to be the union of its intrinsic logics at levels 0 and 1. The superscripts on $\Rightarrow$ could then be dropped. (Again, an analogous observation applies to what follows.) If the level 0 consequence relation is closed under uniform substitution, it would be natural (though by no means mandatory) to suppose that this extends to substitutions of the extended language.

[40] As shown by Barrio et al. (2020).

where all of the formulas of $\Pi$ and $\Gamma$ contain $\Rightarrow^{i-1}$. Let us say that the relation $\vdash^i_w$ such that:

- $A_1, ..., A_m \vdash^i_w B_1, ..., B_n$ iff $w \vDash A_1, ..., A_m \Rightarrow^i B_1, ..., B_n$

is the intrinsic logic of $w$ at level $i$. A world is coherent at this level if the intrinsic logic is an extrinsic logic. @ is coherent at this level as before.

A world is possible if it is coherent at all levels, and its intrinsic logics are the same as those of @ at all levels. If one wishes, one could express all the facts about intrinsic inferences in a single language. We simply take the union of all the languages $\mathfrak{L}^i$, for all finite $i$.

One could iterate the procedure higher up the ordinals if wished, but this is not necessary to express all the individual facts about inferences of finite order. Since premises and conclusions are always finite, if there is a valid inference between formulas of the language in the hierarchy, it will already be expressed by something of the form $A_1, ..., A_m \Rightarrow^i B_1, ..., B_n$, for finite $i$.[41]

## 13.7 Et Cetera

### *13.7.1 The Tarski Rules*

It remains to tie up a few loose ends that have arisen.

Until now, I have assumed that a consequence relation satisfies the familiar Tarski rules of Reflexivity, Monotonicity and Transitivity, since standard logics with world semantics do satisfy these conditions. To a certain extent, these conditions may be relaxed. Let me explain how.

Monotonicity is a defining condition of a deductive logic, and so is not negotiable in the present context. However, there are well known ways of defining deductive logics in which the other two fail.

Essentially, one has to understand what it is to be true at a world in different senses for premises and conclusions. One can do this quite naturally in a many-valued logic. In such logics, some of the values are designated, $D$, and validity is normally defined in terms of the preservation of having values in $D$. However, one may employ different sets of designated values for the premises and conclusions, $D_\Pi$ and $D_\Gamma$. An inference is then valid if whenever all the premises have values in $D_\Pi$ (at a world) some of the consequences have values in $D_\Gamma$ (at that world). If $D_\Pi \subsetneq D_\Gamma$, Transitivity may fail, and if $D_\Gamma \subsetneq D_\Pi$, Reflexivity may fail.

Thus, suppose we deploy the semantics of $K_3/LP$, with values $t$, $i$, and $f$. The matrices for these logics are the same, but in $K_3$ the designated value is just $t$, whilst in $LP$ the designated values are $t$ and $i$.[42] Now, if we take $D_\Pi = \{t\}$, as in $K_3$, and $D_\Gamma = \{t, i\}$, as in $LP$, this gives the logic $ST$ ("Strict/Tolerant"). In this, Transitivity

[41] On iterating levels of meta-inferences into the transfinite, see Scambler (2020).

[42] On $K_3$ and $LP$, see Priest (2008, Chap. 7, esp. Sects. 7.3, 7.4).

fails. Reversing matters gives the logic $TS$ ("Tolerant/Strict"). In this, Reflexivity fails.[43]

With the understanding that truth in the premises and truth in the conclusions are thus understood in different ways, everything above carries over in a straightforward fashion.

### 13.7.2 Non-compact Logics

I have assumed that we are dealing with compact logics, so that we do not have to deal with infinite premise/conclusion sets in any essential way. We could dispense with this assumption. A logic then becomes a relation, $\Pi \vdash \Gamma$, where $\Pi$ and $\Gamma$ are arbitrary sets of formulas of $\mathfrak{L}$. This affects our ability to define the intrinsic logic of a world. One possibility here is to move to an infinitary language, where $\Rightarrow$ can have any number of sentences, finite or infinite, on its left and its right.

A more familiar approach would be obtained by moving to a first-order construction and, in particular, to a language which can refer to sentences, sets thereof, and relations between these things. One can then express the intrinsic logic of a world in the straightforward way, by the condition $w \vDash \langle\Pi\rangle \Rightarrow \langle\Gamma\rangle$, where angle brackets are a name-forming functor, and $\Rightarrow$ is a binary predicate of the familiar kind.[44]

The details of executing this sort of approach are familiar enough, though I shall not pursue them here. The extra complexity would be large, and the gain in understanding for the question at hand virtually nil.

### 13.7.3 Logical Pluralism

Finally, let us turn to the matter of logical pluralism. Logical pluralism is the view that there is, in some sense, more than one correct logic. However, this idea can be cashed out in many different ways, and this is not the place to discuss the matter at length.[45]

The issue is, in fact, irrelevant to the material of Sect. 13.4, since, by assumption, we are dealing with one consequence relation. If there is a plurality of these, each of which is correct, each will determine its own collection of possible/impossible worlds.

---

[43] See, e.g., Cobreros et al. (2013), and Ripley (2013).

[44] And note the following. Suppose that $\mathfrak{L}$ is the language of set theory. Then a model-theoretic definition of validity can be given in that language. Hence, the binary predicate $\Rightarrow$ is already, in effect, present in $\mathfrak{L}$, and does not need to be added to the language.

[45] For some discussion see Priest (2006, Chap. 12).

Matters are different thereafter, since there I have *assumed* that every world (including the actual world) has one intrinsic logic. However, it is not difficult to dispense with this assumption if one is so inclined.

If one is a pluralist, one will have a bunch of possible logics, $\vdash_L$, in play. For each of these, one may have a corresponding connective, $\Rightarrow_L$. We may take each condition of the form $w \vDash A_1, ..., A_m \Rightarrow_L B_1, ..., B_n$ to define an intrinsic logic of $w$. A world is coherent if every intrinsic logic is an extrinsic logic. @ is a coherent world, as before; and now a world is possible if it is coherent and has all the same intrinsic logics as @; it is impossible otherwise.

## 13.8 Conclusion

The considerations of the previous section indicate how the construction of the earlier sections can be enriched and elaborated in various directions. However, such things are not necessary for the main problem of the paper. This was to provide an answer to the question of what an impossible world is. The guiding thought of this essay was that a possible world is one that is closed under logical consequence, and an impossible world is one that is not. As have seen, however, this thought has to be articulated in different ways depending on whether we are dealing with a formal world-semantics for a particular consequence relation, or whether we are dealing with worlds *an sich*, where we had to work to dig out the relevant consequence relation. But in both cases the answer given to the question was independent any assumptions about the correct logic—and many other philosophical matters.

Doubtless, many matters concerning impossible worlds will continue to be debated. I hope that the present paper at least gives us an understanding of what it is we are talking about.

**Acknowledgements** For very helpful written comments on earlier drafts of this paper, I'm grateful to Franz Berto, Allen Hazen, Lloyd Humberstone, Mark Jago, Daniel Nolan, Koji Tanaka, Yale Weiss, Ed Zalta, and two anonymous referees. Talks based on this paper were given in (virtual) seminars at the universities of Utrecht, Melbourne, Padua, Amsterdam, and New York. I'm also grateful to the members of the audiences there for their helpful comments.

## References

Ballarin, R. (2022). *Modern origins of modal logic*. https://plato.stanford.edu/entries/logic-modal-origins/

Barcan Marcus, R. (1995). A backwards look at Quine's animadversions on modalities. *Modalities: Philosophical essays* (pp. 215–232). Oxford: Oxford University Press.

Barrio, E., Pailos, F., & Szmuc, D. (2020). A hierarchy of classical and paraconsistent logics. *Journal of Philosophical Logic, 49*, 93–120.

Berto, F. (2017). Impossible worlds and the logic of imagination. *Erkenntnis, 82*, 1277–1297.

Berto, F., French, R., Priest, G., & Ripley, D. (2018). Williamson on counterpossibles. *Journal of Philosophical Logic, 47*, 693–713.
Berto, F., & Jago, M. (2018). *Impossible worlds*. https://plato.stanford.edu/entries/impossible-worlds/
Berto, F., & Jago, M. (2019). *Impossible worlds*. Oxford: Oxford University Press.
Bjerring, J. C. (2013). Impossible worlds and logical omniscience: An impossibility result. *Synthese, 190*, 2505–2524.
Bjerring, J. C. (2014). On counterpossibles. *Philosophical Studies, 168*, 327–353.
Brogaard, B., & Salerno, J. (2013). Remarks on counterpossibles. *Synthese, 190*, 639–660.
Carroll, J. W. (2020). *Laws of nature*. https://plato.stanford.edu/entries/laws-of-nature/
Cobreros, P., Égré, P., Ripley, D., & van Rooij, R. (2013). Reaching transparent truth. *Mind, 122*, 841–866.
Cresswell, M. (1972). The completeness of S1 and some related systems. *Notre Dame Journal of Formal Logic, 13*(4), 485–496.
Crossley, J., & Humberstone, L. (1977). The logic of 'actually'. *Reports on Mathematical Logic, 8*, 11–29.
Etchemendy, J. (1990). *The concept of logical consequence*. Cambridge: Harvard University Press.
Grandy, R. (1976). Anadic logic and English. *Synthese, 32*, 393–402.
Humberstone, L. (2011). *The connectives*. Cambridge: MIT Press.
Humberstone, L. (2021). Propositional variables occurring exactly once in candidate modal axioms. *Filosofiska Studier, 8*, 27–73.
Jago, M. (2013). Impossible worlds. *Noûs, 49*, 713–728.
Knuuttila, S. (2012). A history of modal traditions. In D. Gabbay, F. J. Pelletier, & J. Woods (Eds.), *Handbook of the history of logic* (Vol. 11). Amsterdam: North Holland.
Krakauer, B. (2013). What are impossible worlds? *Philosophical Studies, 165*, 989–1007.
Kripke, S. (1959). A completeness theorem in modal logic. *Journal of Symbolic Logic, 24*, 1–14.
Kripke, S. (1963). Semantical analysis of modal logic I: Normal modal propositional calculi. *Zeitschrift für mathematische Logik und Grundlagen der Mathematik, 9*, 67–96.
Kripke, S. (1965). Semantical analysis of modal logic II: Non-normal modal propositional calculi. In J. W. Addison, L. Henkin, & A. Tarski (Eds.), *The theory of models* (pp. 202–220). Amsterdam: North Holland.
McKinsey, J., & Tarski, A. (1944). The algebra of topology. *Annals of Mathematics, 45*, 141–191.
McKinsey, J., & Tarski, A. (1946). On closed elements in closure algebras. *Annals of Mathematics, 47*, 122–162.
Menzel, C. (2016). *Possible worlds*. https://plato.stanford.edu/entries/possible-worlds/
Nolan, D. (2013). Impossible worlds. *Philosophy Compass, 8*, 360–372.
Nolan, D. (2014). Hyperintensional metaphysics. *Philosophical Studies, 171*, 149–160.
Priest, G. (1992). What is a non-normal world? *Logique et Analyse, 35*, 291–302.
Priest, G. (1997a). Editor's introduction. *Notre Dame Journal of Formal Logic*, *38*(4), 482–487. (In Priest (1997b).)
Priest, G. (1997b). Special issue on impossible worlds. *Notre Dame Journal of Formal Logic*, *38*(4).
Priest, G. (2005). *Towards non-being*. Oxford: Oxford University Press. (2nd edn published in 2016.)
Priest, G. (2006). *Doubt truth to be a liar*. Oxford: Oxford University Press.
Priest, G. (2008). *Introduction to non-classical logic*. Cambridge: Cambridge University Press.
Priest, G. (2021). Metaphysical necessity: A skeptical perspective. *Synthese, 198*, 1873–1885.
Ripley, D. (2013). Paradoxes and failures of cut. *Australasian Journal of Philosophy, 91*, 139–164.
Routley, R., Plumwood, V., Meyer, R. K., & Brady, R. (1982). *Relevant logics and their rivals* (Vol. 1). Atascadero: Ridgeview.
Scambler, C. (2020). Transfinite meta-inferences. *Journal of Philosophical Logic, 49*, 1079–1089.
Sillari, G. (2008). Quantified logic of awareness and impossible possible worlds. *Review of Symbolic Logic, 4*, 514–529.

Stuart, M., McLoone, B., & Grützner, C. (2023). Counterpossibles in science: An experimental study. *Synthese, 201*, 1–20.
Tanaka, K. (2018). Logically impossible worlds. *Australasian Journal of Logic, 15*(2), 489–497.
Taylor, B., & Hazen, A. (1992). Flexibly structured predication. *Logique et Analyse, 35*(139/140), 375–393.
Weber, Z., & Omori, H. (2019). Observations on the trivial world. *Erkenntnis, 84*, 975–994.
Yagisawa, T. (2010). *Worlds and individuals, possible and otherwise*. Oxford: Oxford University Press.
Zalta, E. (1997). A classically-based theory of impossible worlds. *Notre Dame Journal of Formal Logic, 38*, 640–660.

# Chapter 14
# Substructural Negations as Normal Modal Operators

**Heinrich Wansing**

**Abstract** A theory of substructural negations as impossibility and as unnecessity based on bi-intuitionistic logic, also known as Heyting-Brouwer logic, has been developed by Takuro Onishi. He notes two problems for that theory and offers the identification of the two negations as a solution to both problems. The first problem is the lack of a structural rule corresponding with double negation elimination for negation as impossibility, DNE, and the second problem is a lack of correspondence between certain sequents and a characterizing frame property. While the identification of negation as impossibility and negation as unnecessity does solve Onishi's problems, in general it nevertheless seems desirable to keep the two notions separate. The present paper addresses the first problem by introducing a reformulation of Onishi's display sequent calculus in a language of sequents that incorporates Boolean negation. Instead of identifying negation as impossibility and negation as unnecessity, a notion of weak frame correspondence is defined. It is observed that DNE weakly corresponds with a certain frame property, and two structural sequent rules are presented that together also weakly correspond with that frame property and allow one to derive DNE. Moreover, the reformulated display calculus has an independent motivation by considerations on proof-theoretic semantics.

**Keywords** Negation · Heyting-Brouwer logic · Negation as a modal operator · Display calculi · Correspondence · Proof-theoretic semantics

## 14.1 Introduction

This is a paper on the theory of negation as a normal modal operator of impossibility or unnecessity, a theory that would not exist without the seminal work of Kripke (1959, 1963a, b, c, 1965) on modal logic.[1] The study of negation as incompatibility

[1] I dedicate this paper to professor Saul Aaron Kripke.

H. Wansing (✉)
Department of Philosophy I, Ruhr University Bochum, Bochum, Germany
e-mail: Heinrich.Wansing@rub.de

© The Author(s), under exclusive license to Springer Nature Switzerland AG 2024

Y. Weiss and R. Birman (eds.), *Saul Kripke on Modal Logic*, Outstanding Contributions to Logic 30, https://doi.org/10.1007/978-3-031-57635-5_14

has a long tradition and dates back to work by Birkhoff and von Neumann (1936) on negation in quantum logic, later investigated by Goldblatt (1974). Further developments are due to Vakarelov (1977, 1989a, b), Došen (1984, 1986, 1999), Dunn (1993, 1996, 1999), Dunn and Zhou (2005), Restall (1999), and Mares (2004). The conception of negation as a normal modal operator has also attracted a lot of attention and discussion more recently in the work of Berto (2014), Onishi (2015), Lahav et al. (2017), Drobyshevich (2018), Berto and Restall (2019), Horn and Wansing (2015), and Odintsov and Wansing (2021).

In Horn and Wansing (2015) the question is posed whether sentential negation indeed *is* a normal modal operator of impossibility or unnecessity, and it is remarked that in the opinion of Berto (2014), the meaning of negation is grounded in the notion of compatibility, together with its opposite concept of incompatibility. Since Berto takes compatibility and incompatibility to be symmetric relations, negation as impossibility, $\rhd$, not only satisfies the contraposition rule, $A \vdash B / \rhd B \vdash \rhd A$, but also the double negation introduction rule, $A \vdash \rhd \rhd A$. Berto therefore holds that no unary connective can be justifiably called a negation if it does not satisfy contraposition and double negation introduction. Berto and Restall (2019, p. 1119) go one step further and explain that "because incompat*ibility* is modal, negation is a modal operator as well." However, this view is contentious. In light of non-contraposable negations in prominent three-valued logics such as Łukasiewicz's system Ł3, Kleene's strong three-valued logic, K3, Asenjo and Priest's Logic of Paradox, LP, (see, e.g., Priest, 2008), and four-valued logics such as Nelson's constructive logic N4 and its connexive variant C (see, e.g., Almukdad & Nelson, 1984, Wansing, 2005, 2006, Omori & Wansing, 2020), Berto and Restall's claim that sentential negation *is* a normal modal operator may be doubted, and it has been carefully argued by De and Omori (2018) that there is more to negation than modality.

The present paper examines Takuro Onishi's 2015 theory of negation as impossibility and negation as unnecessity over bi-intuitionistic logic, also known as Heyting-Brouwer logic (see Goré, 2000, Wansing, 2016, Goré, 2020, and references therein). This approach is remarkable because Onishi presents negation as impossibility and negation as unnecessity as *substructural* negations that can be obtained from classical negation by giving up structural sequent rules in a suitable display sequent calculus. Onishi's work has been extended by Drobyshevich (2018), who presents display calculi for non-normal versions of impossibility and unnecessity operators and establishes connections between Onishi's and Dimiter Vakarelov's research on negation. Moreover, Drobyshevich obtains decidability results for a large number of display calculi for logics with negative modal operators, including Onishi's display calculus.[2]

Building on Nuel Belnap's seminal (1982) paper in which he introduced display logic, in Wansing (1998) the modal display calculus has been suggested as a framework for a proof-theoretic semantics of normal modal operators together with a checklist of conditions on the shape of introduction rules in a sequent calculus.

[2] See Ciabattoni and Ramanayake (2016) for a study of what can and cannot be expressed by structural rules in display sequent calculi.

These conditions contribute to making sure that the introduction rules can be seen as defining the meaning of logical operations inferentially (see also Poggiolesi, 2010; Parisi, 2022). Moreover, in Wansing (1994) an idea of Došen (1988, p. 353) is taken up, namely that in defining non-classical logics,"the rules for the logical operations are never changed: all changes are made in the structural rules." If the left and right introduction rules of the logical operations in a suitable sequent calculus are kept constant, the basic proof-theoretic meaning of these operations is fixed and a modular account of different readings of the logical constants is obtained by varying the structural rules of the calculus, the rules that do not exhibit any logical operation of the object language. This conception has been called 'Došen's principle' by Wansing (1994). Onishi's theory thus implements Došen's principle for negation as a normal modal operator of impossibility or unnecessity.

One of the meaning-theoretically significant conditions on sequent rules for logical operations is *separation* (or purity): a sequent rule for introducing a logical operation *op* in antecedent or succedent position of a sequent should exhibit only one logical operation because otherwise it specifies the meaning of *op* in combination with other logical operations, which would amount to a kind of semantic holism. The reformulated version of Onishi's display calculus for negation as impossibility and negation as unnecessity satisfies separation not only with respect to the sequent rules for the logical operations from the object language but also for the structural rules governing the structure connectives from the language of display sequents, including the so-called *display rules*. Besides addressing Onishi's first problem, reformulating his display calculus has an independent motivation, namely making the meaning of the structural operations as given with the structural sequent rules independent of each other. Note that the structure connectives still have a context-sensitive reading, depending on whether they appear in antecedent or succedent position. The new introduction rules for the negations as modal operators *do* exhibit more than one structure connective, but this only reveals that the meaning of these negations is given in terms of the combination of two structural operations, namely an operation of context shift and a context-dependent marking as a modal operator.

The paper is structured as follows. First, in Sect. 14.2, Onishi's theory of substructural negations is recalled. Next, in Sect. 14.3, Onishi's two problems are explained and the solution he offers to solve the two problems in one go. Another approach to addressing Onishi's first problem is presented in Sect. 14.4. It proceeds by re-displaying Onishi's sequent calculus with the help of a language of structures that comprises a counterpart of classical negation. This move makes it possible to state structural sequent rules that correspond with double negation elimination for negation as impossibility in a weak sense. As a by-product, we obtain a new cut-free display sequent calculus for propositional Heyting-Brouwer logic. The paper ends with some brief concluding remarks in Sect. 14.5.

## 14.2 Onishi's Theory of Substructural Negations

The framework of substructural negations developed by Onishi (2015) starts from bi-intuitionistic logic. Onishi introduces four negations as normal modal operators, namely forward-looking impossibility, $\triangleright$, and backward-looking impossibility, $\triangleleft$, as well as forward-looking unnecessity, $\blacktriangleright$, and backward-looking unnecessity, $\blacktriangleleft$. In their relational semantics, both pairs come with their own binary accessibility relation, namely $\frown$ as the compatibility relation used to interpret negation as impossibility, and $\smile$ as a separate relation of "exhaustiveness" used to interpret negation as unnecessity.

The propositional language $\mathcal{L}$ extends the language of bi-intuitionistic logic by the four negations as modal operators and is defined in Backus–Naur form as follows:

atomic formulas: $p \in Atom$
formulas: $A \in Form_{\mathcal{L}}(Atom)$
$A ::= p \mid (A \wedge A) \mid (A \vee A) \mid (A \rightarrow A) \mid (A \leftarrow A) \mid \triangleright A \mid \triangleleft A \mid \blacktriangleright A \mid \blacktriangleleft A.$

The truth constant $\top$ and the falsity constant $\bot$ are defined as $(p \rightarrow p)$, respectively $(p \leftarrow p)$, for some atomic formula $p$. Then intuitionistic negation, $\neg$, and co-negation, $-$, are defined as follows: $\neg A := (A \rightarrow \bot)$; $-A := (\top \leftarrow A)$.[3]

**Definition 14.2.1** A $BiN$-frame is a structure $(W, \leq, \frown, \smile)$, where $W$ is a non-empty set of states, $\leq$ is a reflexive and transitive relation (an information order) on $W$, and $\frown$ ("compatibility"), and $\smile$ ("exhaustiveness") are binary accessibility relations on $W$ satisfying the following conditions:

(i) $\leq \circ \frown = \frown = \frown \circ \geq$ and (ii) $\geq \circ \smile = \smile = \smile \circ \leq$,

where $\geq$ is the inverse of $\leq$, and $\circ$ is the composition of relations. A $BiN$-model based on a $BiN$-frame $(W, \leq, \frown, \smile)$ is a structure $(W, \leq, \frown, \smile, \models)$, where $\models \subseteq W \times Atom$ and heredity (persistence) for atomic formulas holds: for every $x, y \in W$ and $p \in Atom$, if $x \models p$ and $x \leq y$, then $y \models p$. The relation $\models$ is extended to a relation between states and arbitrary $\mathcal{L}$-formulas as follows:

$$
\begin{array}{ll}
x \models (A \wedge B) & \text{iff } x \models A \text{ and } x \models B \\
x \models (A \vee B) & \text{iff } x \models A \text{ or } x \models B \\
x \models (A \rightarrow B) & \text{iff } \forall y(x \leq y \text{ implies } (y \not\models A \text{ or } y \models B)) \\
x \models (A \leftarrow B) & \text{iff } \exists y(y \leq x,\ y \models A, \text{ and } y \not\models B) \\
x \models \triangleright A & \text{iff } \forall y(x \frown y \text{ implies } y \not\models A) \\
x \models \triangleleft A & \text{iff } \forall y(y \frown x \text{ implies } y \not\models A) \\
x \models \blacktriangleright A & \text{iff } \exists y(x \smile y \text{ and } y \not\models A \\
x \models \blacktriangleleft A & \text{iff } \exists y(y \smile x \text{ and } y \not\models A.
\end{array}
$$

[3] Here I use the symbol '$\leftarrow$' for co-implication because it is used by Onishi (2015) and Drobyshevich (2018). At other places I have mostly used the symbol '$\multimap$' that was introduced in Goré (2000).

**Table 14.1** The understanding of structure connectives as logical operations

| | **I** | ; | > | ♯ | ♭ |
|---|---|---|---|---|---|
| in antecedent position | $\top$ | $\wedge$ | $\leftarrow$ | $\blacktriangleright$ | $\blacktriangleleft$ |
| in succedent position | $\bot$ | $\vee$ | $\rightarrow$ | $\triangleright$ | $\triangleleft$ |

Note that we can define a forward-looking necessity operator $\Box^+$ with respect to $\leq$ and a backward-looking possibility operator $\Diamond^-$ with respect to $\leq$ by the following evaluation clauses:

$$x \models \Box^+ A \text{ iff } x \models \top \rightarrow A, \qquad x \models \Diamond^- A \text{ iff } x \models \top \leftarrow A.$$

**Definition 14.2.2** An $\mathcal{L}$-formula $A$ is true in a $Bi$N-model $(W, \leq, \frown, \smile, \models)$ if $x \models A$ for every $x \in W$; and $A$ is valid on a $Bi$N-frame $(W, \leq, \frown, \smile)$ if $A$ is true in every model based on $(W, \leq, \frown, \smile)$. Moreover, $A$ is valid with respect to a class of models (frames) if $A$ is true in every model (valid on every frame) from that class. The logic semantically defined as the set of all $\mathcal{L}$-formulas valid with respect to the class of all $Bi$N-frames is called $Bi$N.

Heredity holds for arbitrary $\mathcal{L}$-formulas; the condition (i), respectively, (ii) of a $Bi$N-frame makes sure that it holds for formulas prefixed by $\triangleright$, $\triangleleft$, respectively $\blacktriangleright$, and $\blacktriangleleft$.

Two different display calculi for bi-intuitionistic logic have been introduced in Goré (2000) and Wansing (2008), see also Kamide and Wansing (2015) and Ciabattoni et al. (2014). Onishi (2015) defines a display calculus $\delta Bi$N for $Bi$N that builds on Rajeev Goré's version from Goré (2000).

**Definition 14.2.3** The language of structures of Onishi's display sequent calculus $\delta Bi$N based on $\mathcal{L}$ is defined as follows:

formulas: $A \in \text{Form}_{\mathcal{L}}(\text{Atom})$
structures: $X \in \text{Struc}(\text{Form}_{\mathcal{L}}(\text{Atom}))$
$X ::= A \mid \mathbf{I} \mid \sharp X \mid \flat X \mid (X; X) \mid (X > X)$.[4]
An o-sequent is an expression of the form $X \vdash Y$, where $X$ and $Y$ are structures from $\text{Struc}(\text{Form}_{\mathcal{L}}(\text{Atom}))$.

The constant **I**, the unary structure connectives ♯ and ♭, and the binary structure connectives ; and > have a context sensitive reading, depending on whether they appear in antecedent or succedent position of an o-sequent. The understanding of the structure connectives in terms of logical operations from $\mathcal{L}$ is shown in Table 14.1, and can be made precise by the translation $\tau$ from o-sequents to $\mathcal{L}$-formulas given in Definition 14.2.4.

[4] In what follows, outer brackets in formulas and structures will sometimes be omitted.

**Definition 14.2.4** The translation $\tau$ from o-sequents to $\mathcal{L}$-formulas is defined by setting $\tau(X \vdash Y) = \tau_1(X) \rightarrow \tau_2(Y)$, where

$$\begin{array}{rlrl}
\tau_1(A) &= A & \tau_2(A) &= A \\
\tau_1(\mathbf{I}) &= \top & \tau_2(\mathbf{I}) &= \bot \\
\tau_1(X; Y) &= \tau_1(X) \wedge \tau_1(Y) & \tau_2(X; Y) &= \tau_2(X) \vee \tau_2(Y) \\
\tau_1(X > Y) &= \tau_1(X) \leftarrow \tau_2(Y) & \tau_2(X > Y) &= \tau_1(X) \rightarrow \tau_2(Y) \\
\tau_1(\sharp X) &= \blacktriangleright \tau_2(X) & \tau_2(\sharp X) &= \triangleright \tau_1(X) \\
\tau_1(\flat X) &= \blacktriangleleft \tau_2(X) & \tau_2(\flat X) &= \triangleleft \tau_1(X).
\end{array}$$

**Definition 14.2.5** The display sequent calculus $\delta Bi\mathrm{N}$ is defined as the set consisting of the following axiom and rules, where double-lines indicate that a rule can be applied also in inverted-form:

Identity and cut

$$\frac{}{p \vdash p}\ (id) \qquad \frac{X \vdash A \quad A \vdash Y}{X \vdash Y}\ (cut)$$

Display rules

$$\frac{X; Y \vdash Z}{\dfrac{X \vdash Y > Z}{Y; X \vdash Z}} \qquad \frac{X \vdash Y; Z}{\dfrac{X > Y \vdash Z}{X \vdash Z; Y}} \qquad \frac{\sharp X \vdash Y}{\flat Y \vdash X} \qquad \frac{X \vdash \sharp Y}{Y \vdash \flat X}$$

Structural rules

$$\frac{X; \mathbf{I} \vdash Y}{X \vdash Y} \qquad \frac{X \vdash Y; \mathbf{I}}{X \vdash Y} \qquad \frac{(X; Y); Z \vdash W}{X; (Y; Z) \vdash W} \qquad \frac{X \vdash Y; (Z; W)}{X \vdash (Y; Z); W}$$

$$\frac{X \vdash Y}{X; Z \vdash Y} \qquad \frac{X \vdash Y}{X \vdash Y; Z} \qquad \frac{X; X \vdash Y}{X \vdash Y} \qquad \frac{X \vdash Y; Y}{X \vdash Y}$$

Operational rules

$$\frac{A; B \vdash Y}{A \wedge B \vdash Y} \qquad \frac{X \vdash A \quad Y \vdash B}{X; Y \vdash A \wedge B} \qquad \frac{A \vdash X \quad B \vdash Y}{A \vee B \vdash X; Y} \qquad \frac{X \vdash A; B}{X \vdash A \vee B}$$

$$\frac{X \vdash A \quad B \vdash Y}{A \rightarrow B \vdash X > Y} \qquad \frac{X \vdash A > B}{X \vdash A \rightarrow B} \qquad \frac{A > B \vdash Y}{A \leftarrow B \vdash Y} \qquad \frac{X \vdash A \quad B \vdash Y}{X > Y \vdash A \leftarrow B}$$

$$\frac{X \vdash A}{\triangleright A \vdash \sharp X} \qquad \frac{X \vdash \sharp A}{X \vdash \triangleright A} \qquad \frac{X \vdash A}{\triangleleft A \vdash \flat X} \qquad \frac{X \vdash \flat A}{X \vdash \triangleleft A}$$

$$\frac{\sharp A \vdash Y}{\blacktriangleright A \vdash Y} \qquad \frac{A \vdash Y}{\sharp Y \vdash \blacktriangleright A} \qquad \frac{\flat A \vdash Y}{\blacktriangleleft A \vdash Y} \qquad \frac{A \vdash Y}{\flat Y \vdash \blacktriangleleft A}$$

Note that for any $\mathcal{L}$-formula $A$, the o-sequent $A \vdash A$ is derivable in $\delta Bi\mathrm{N}$.

**Definition 14.2.6** (*Antecedent and succedent parts*) If $s = X \vdash Y$ is an o-sequent, then $X$ is an antecedent part (AP) of $s$ and $Y$ is a succedent part (SP) of $s$. Moreover:

- If $W; Z$ is an AP (SP) of $s$, then so are $W$ and $Z$;
- If $W > Z$ is an AP or an SP in $X \vdash Y$, then $W$ is an AP and $Z$ is an SP of $s$;
- If $\sharp W$ is an AP (SP) of $s$, then $W$ is an SP (AP) of $s$;
- If $\flat W$ is an AP (SP) of $s$ then $W$ is an SP (AP) of $s$.

**Definition 14.2.7** Two o-sequents are display equivalent (or structurally equivalent) if they are derivable from each other using only display rules.

**Theorem 14.2.8** (Display property) *Let $Z$ be a substructure of the o-sequent $s = X \vdash Y$. Then,*

- *If $Z$ is an AP of $s$, then there is an o-sequent $Z \vdash W$ that is display equivalent to $s$.*
- *If $Z$ is an SP of $s$, then there is an o-sequent $W \vdash Z$ that is display equivalent to $s$.*

In view of the display property, the left and right introduction rules for the logical operations can be stated with the principal formulas as the entire antecedent, respectively succedent of the conclusion o-sequent. Moreover, since $\delta BiN$ is a proper display calculus (cf. Belnap, 1982; Wansing, 1998), Belnap's general cut-elimination theorem applies.

**Theorem 14.2.9** (Cut elimination) *If an o-sequent $X \vdash Y$ is derivable in $\delta BiN$, it is derivable without applications of (cut).*

**Definition 14.2.10** An o-sequent $X \vdash Y$ is valid with respect to a class of frames iff the $\mathcal{L}$- formula $\tau(X \vdash Y)$ is valid with respect to that class.

**Theorem 14.2.11** (Onishi, 2015) *An o-sequent $X \vdash Y$ is derivable in $\delta BiN$ if and only if $X \vdash Y$ is valid with respect to the class $\mathcal{BIN}$ of all $BiN$-frames.*

Soundness is proved by induction on derivations in $\delta BiN$. The canonical model construction used to prove Theorem 14.2.11 is presented in the Appendix.

The following o-sequents are valid on a $BiN$-frame if and only if the first-order conditions given with them are satisfied. In each but one case a corresponding structural sequent rule is presented in linear notation (i.e., a $BiN$-frame satisfies the frame condition if and only if the rule preserves validity on the frame):

| | | |
|---|---|---|
| DNI | $A \vdash \rhd \rhd A$<br>$X \vdash \sharp Y / Y \vdash \sharp X$ | $\forall x \forall y (x \frown y \Rightarrow y \frown x)$ |
| Con | $(A \to B) \vdash (\rhd B \to \rhd A)$<br>$X \vdash \sharp (Y; Z) / X \vdash Y > \sharp Z$ | $\forall x \forall y (x \frown y \Rightarrow \exists z (x \leq z \& y \leq z \& x \frown z))$ |
| LNC | $(A \wedge \rhd A) \vdash B$<br>$X \vdash \sharp Y / X \vdash Y > \mathbf{I}$ | $\forall x\, x \frown x$ |
| $\rhd\top\bot$ | $\rhd\top \vdash \bot$<br>$X \vdash \sharp \mathbf{I} / X \vdash \mathbf{I}$ | $\forall x \exists y\, x \frown y$ |
| LEM | $B \vdash A \vee \rhd A$<br>$X; Y \vdash Z / X \vdash \sharp Y; Z$ | $\forall x \forall y (x \frown y \Rightarrow y \leq x)$ |
| FDM | $\rhd (A \wedge B) \vdash (\rhd A \vee \rhd B)$<br>$X \vdash \sharp (Y; Z) / X \vdash \sharp Y; \sharp Z$ | $\forall x \forall y \forall z ((x \frown y \& x \frown z) \Rightarrow$<br>$\exists w (y \leq w \& z \leq w \& x \frown w))$ |
| DNE | $\rhd \rhd A \vdash A$ | [DNE] $\forall x \exists y (x \frown y \& \forall z (y \frown z \Rightarrow z \leq x))$ |

The listed frame conditions are canonical. If $\delta Bi\mathrm{N}$ is extended by one of the axiomatic sequents, then the frame of the canonical model for that extension satisfies the corresponding condition. Also, cut-elimination holds for extensions of $\delta Bi\mathrm{N}$ with any combination of the above structural sequent rules because the resulting systems are proper display calculi.

From the point of view of proof-theoretic semantics, the correspondences with structural rules are appealing for it may be held that the above introduction rules of the four negations once and for all fix their meaning, so that variations of meaning can be spelled out in purely structural terms. The emerging 'kite' of negations as impossibility is shown in Fig. 14.1.[5]

## 14.3 Onishi's Problems and His Solution

Whereas the first six of the above o-sequents correspond with the listed purely structural sequent rules, for double negation elimination it is not clear whether there exists a corresponding structural rule. There is thus a lack of a structural rule. This is the first problem highlighted by Onishi. The second problem is a lack of correspondence. The extension of $\delta Bi\mathrm{N}$ obtained by adding the o-sequents $\rhd\top \vdash \bot$ and $\rhd (A \wedge B) \vdash (\rhd A \vee \rhd B)$ (the logic of Ockham negation) is complete with respect to the class of all "star-crossed" $Bi\mathrm{N}$-frames, i.e., frames satisfying the following condition:

$$(*) \quad \forall x \exists y (x \frown y \& \forall w (x \frown w \Rightarrow y \geq w))$$

[5] The diagram of negation operators presented in Dunn (1993, 1996, 1999) has the shape of a kite. Although some diagrams displaying later modifications of that kite do not really resemble a kite, the terminology was kept.

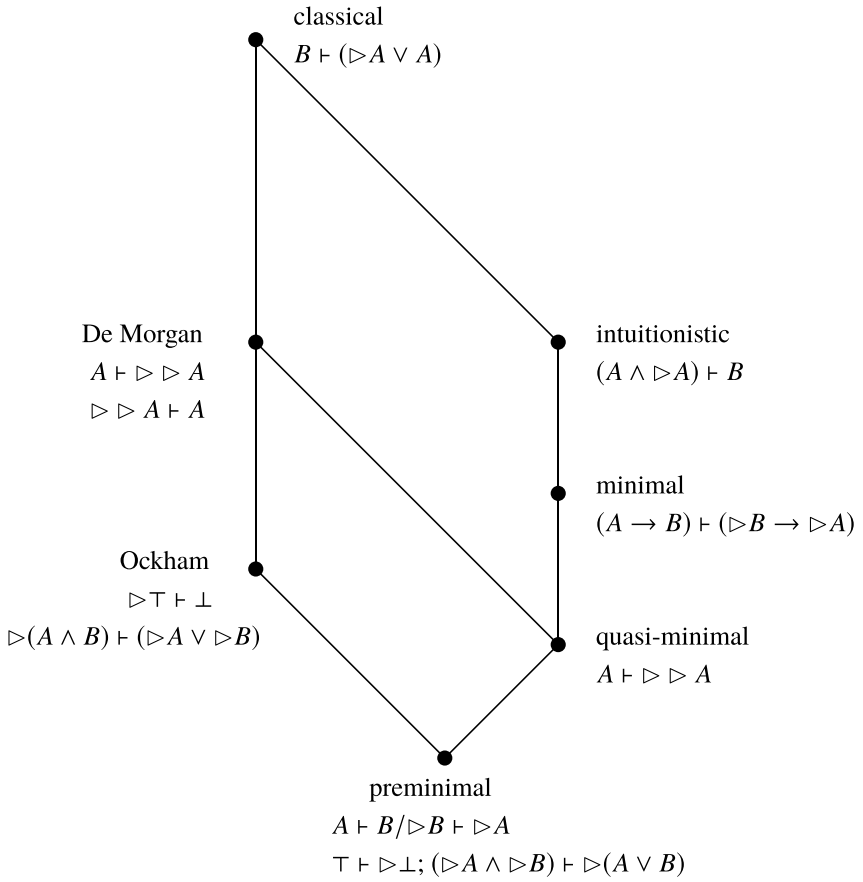

classical
$B \vdash (\triangleright A \vee A)$
De Morgan
$A \vdash \triangleright \triangleright A$
$\triangleright \triangleright A \vdash A$
intuitionistic
$(A \wedge \triangleright A) \vdash B$
minimal
$(A \rightarrow B) \vdash (\triangleright B \rightarrow \triangleright A)$
Ockham
$\triangleright \top \vdash \bot$
$\triangleright(A \wedge B) \vdash (\triangleright A \vee \triangleright B)$
quasi-minimal
$A \vdash \triangleright \triangleright A$
preminimal
$A \vdash B / \triangleright B \vdash \triangleright A$
$\top \vdash \triangleright \bot; (\triangleright A \wedge \triangleright B) \vdash \triangleright(A \vee B)$

**Fig. 14.1** The kite of negation as impossibility

The condition (*), however, does not correspond with the combination of the two o-sequents, as Onishi shows that there is a frame that satisfies the frame conditions corresponding to them but fails to satisfy (*).

Onishi solves both problems by identifying the negations $\triangleright$ and $\blacktriangleright$. He defines a mapping, $(\cdot)^t$, (notation adjusted) on $\mathcal{L}$-formulas as follows:

$$\begin{array}{rclcrcl} (A \wedge B)^t & = & (A^t \vee B^t) & & (A \vee B)^t & = & (A^t \wedge B^t) \\ (A \rightarrow B)^t & = & (B^t \leftarrow A^t) & & (A \leftarrow B)^t & = & (B^t \rightarrow A^t) \\ (\triangleright A)^t & = & \blacktriangleright A^t & & (\triangleleft A)^t & = & \blacktriangleleft A^t \\ (\blacktriangleright A)^t & = & \triangleright A^t & & (\blacktriangleleft A)^t & = & \triangleleft A^t \end{array}$$

This translation is extended to o-sequents and to the first-order language of conditions on frames (the latter by replacing every occurrence of $\frown$ with an occurrence of $\smile$, and vice versa, and every occurrence of $\leq$ with an occurrence of $\geq$, and vice versa),

so as to obtain dual frame correspondences for o-sequents expressing properties of $\blacktriangleright$ and a dual kite of negations as unnecessity, charting negations dual to the ones from the kite of negations as impossibility. Moreover, Onishi notes that the dual Ockham, dual de Morgan, and dual classical negations are just the Ockham, de Morgan, and classical negations expressed by unnecessity.

Since the characteristic principles of preminimal negation as impossibility are translated as the characteristic principles of negation as unnecessity, and vice versa, the identification of $\triangleright$ and $\blacktriangleright$ results in Ockham negation. The o-sequents that express the identification of negation as impossibility and negation as unnecessity, $\triangleright A \vdash \blacktriangleright A$ and $\blacktriangleright A \vdash \triangleright A$, correspond with the following frame conditions and structural sequent rules:

$$\begin{array}{lll} \triangleright A \vdash \blacktriangleright A & \forall x \exists y (x \frown y \& x \smile y) & \flat X \vdash \flat Y \,/\, Y \vdash X \\ \blacktriangleright A \vdash \triangleright A & \forall x \forall y \forall z ((x \frown y \& x \smile z) \Rightarrow y \leq z) & X \vdash Y \,/\, \sharp Y \vdash \sharp X \end{array}$$

With the identification of negation as impossibility and negation as unnecessity we are led to the united kite of negations shown in Fig. 14.2.

Note that in the presence of the o-sequents for identifying negation as impossibility and negation as unnecessity and their frame conditions, every negation from the united kite of negations emerges as substructural because both double negation elimination for $\blacktriangleright$ and double negation introduction for $\triangleright$ do correspond with a structural sequent rule. This solves Onishi's first problem.

Moreover, Onishi points out that the two frame conditions corresponding with the sequents $\triangleright A \vdash \blacktriangleright A$ and $\blacktriangleright A \vdash \triangleright A$ are equivalent to a condition that is stronger than (*), and requires for every state $x$ the existence of a state $x^*$ that is not only maximal among the states compatible with $x$ but also minimal among the states that are jointly exhaustive with $x$, so that $x \models \triangleright A$ iff $x^* \not\models A$ iff $x \models \blacktriangleright A$:

$$\forall x \exists y ((x \frown y \& x \smile y) \& \forall z (x \frown z \Rightarrow z \leq y) \& \forall w (x \smile w \Rightarrow y \leq w))$$

Since the frame conditions for the sequents $\triangleright A \vdash \blacktriangleright A$ and $\blacktriangleright A \vdash \triangleright A$ are canonical, the problem is solved (although still no sequent or sequent rule has been found that corresponds with (*)).

## 14.4 Another Approach to Onishi's First Problem

We will re-display Onishi's sequent calculus by using a language of structures that comprises a counterpart of classical negation. Sequents are then understood in terms of a translation into the language $\mathcal{L}^{\neg}$ that extends $\mathcal{L}$ by Boolean negation, $\neg$. Note that the language of $Bi\mathrm{N}$ is still $\mathcal{L}$; there will be no introduction rules for $\neg$.

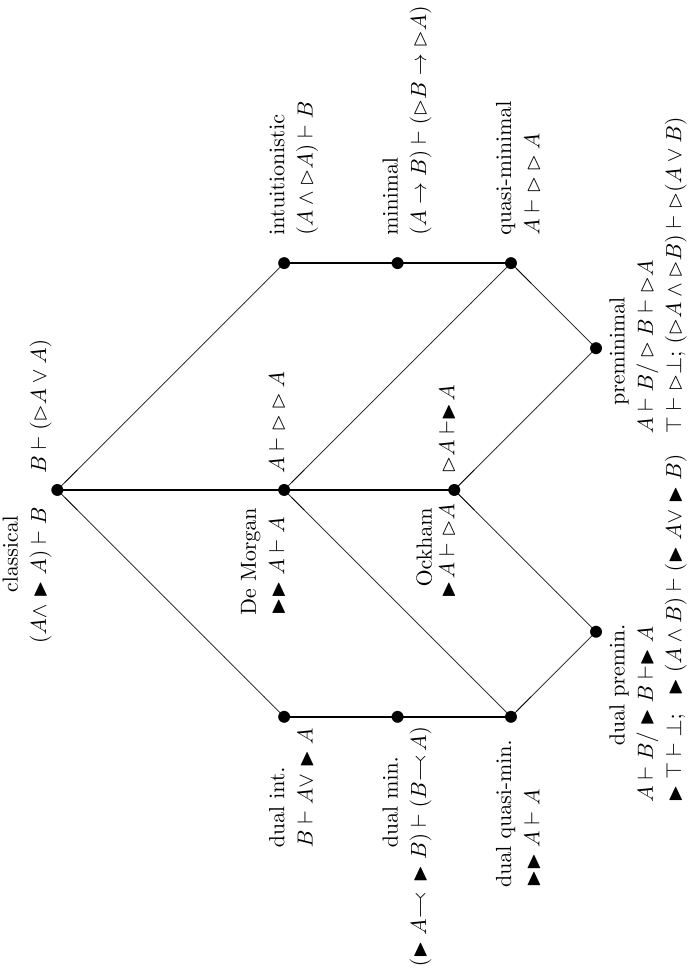



**Fig. 14.2** The united kite of negations

The propositional language $\mathcal{L}^{\neg}$ is thus defined as follows:

atomic formulas: $p \in Atom$
formulas: $A \in Form_{\mathcal{L}^{\neg}}(Atom)$
$A ::= p \mid (A \wedge A) \mid (A \vee A) \mid (A \to A) \mid (A \leftarrow A) \mid \rhd A \mid \lhd A \mid \blacktriangleright A \mid \blacktriangleleft A \mid \neg A.$

The presence of classical negation allows one to define a forward-looking necessity operator, $\Box^{+}$, a backward-looking possibility operator, $\Diamond^{-}$, a forward-looking possibility operator, $\blacklozenge^{+}$, and a backward-looking necessity operator, $\blacksquare^{-}$, as follows:

$$\Box^{+}A := \rhd\neg A, \qquad \Diamond^{-}A := \neg \lhd A, \qquad \blacklozenge^{+}A := \blacktriangleright\neg A, \qquad \blacksquare^{-}A := \neg\blacktriangleleft A.$$

Also, we can define classical implication, $\rightarrow_c$, and classical co-implication, $\leftarrow_c$, as follows:

$$(A \rightarrow_c B) := (\neg A \vee B), \qquad (A \leftarrow_c B) := (A \wedge \neg B).$$

**Definition 14.4.1** The notion of a $BiN$-model is defined as before, except that the relation $\models$ is extended to a relation between states and arbitrary $\mathcal{L}^{\neg}$-formulas by determining that $x \models \neg A$ iff $x \not\models A$.

Two $\mathcal{L}^{\neg}$-formulas $A$ and $B$ are said to be logically equivalent iff $((A \rightarrow B) \wedge (B \rightarrow A))$ is valid with respect to the class of all $BiN$-frames. Note that the addition of classical negation to the semantically defined logic $BiN$ is conservative; Peirce's Law, $((A \rightarrow B) \rightarrow A) \rightarrow A$, and the negation-free version of the Law of Excluded Middle, $A \vee (A \rightarrow B)$, are still refutable. Obviously, heredity does not hold for arbitrary $\mathcal{L}^{\neg}$-formulas (but it continues to hold for $\mathcal{L}^{\neg}$-formulas that are logically equivalent to $\mathcal{L}$-formulas). For the language of display sequents, I will build on the display calculus for bi-intuitionistic logic from Wansing (2008)[6] because there the display rules, like the other structural rules, are already separated, i.e., they exhibit only one structure connective. Moreover, I will build on Goré's display calculus for intuitionistic logic in Goré (1995), which exploits the faithful embedding of intuitionistic logic into the modal logic S4, and the faithful embedding of Heyting-Brouwer logic into temporal S4, see Łukowski (1996) and Wolter (1998).

**Definition 14.4.2** The language of structures of the sequent calculus $dBiN$, which we are about to present, is defined over the set of all $\mathcal{L}$-formulas as follows:

formulas: $A \in \mathrm{Form}_{\mathcal{L}}(\mathrm{Atom})$
structures: $X \in \mathrm{Struc}^*(\mathrm{Form}_{\mathcal{L}})$
$X ::= A \mid \mathbf{I} \mid *X \mid \odot X \mid \circledcirc X \mid \circledast X \mid (X \circ X) \mid (X \bullet X)$.

A b-sequent is an expression of the form $X \vdash Y$, where $X$ and $Y$ are structures from $\mathrm{Struc}^*(\mathrm{Form}_{\mathcal{L}})$.

The reading of the structure connectives as operations from $\mathcal{L}^{\neg}$ is as stated in Table 14.2, and it is made precise by the translation $\sigma$ from Definition 14.4.3. The language of structures thus contains the connectives of *classical* Priorean temporal logic with backward- and forward-looking modalities with respect to the binary relations $\frown$, $\smile$, and $\leq$ from $BiN$-frames. The use of classical propositional logic as the non-modal base logic is not worrying because the meta-language used to define $BiN$ semantically is classical. The introduction rules for $\rightarrow$ on the right and for $\leftarrow$ on the left side of the turnstile make use of structure connectives that are understood as the modal operators $\Box^+$ and $\Diamond^-$ associated with $\leq$.

**Definition 14.4.3** The translation $\sigma$ from b-sequents to $\mathcal{L}^{\neg}$-formulas is defined by setting

[6] We will use one structure connective that in its context-sensitive reading varies between conjunction and Boolean implication, and another structure connective that in its context-sensitive reading varies between disjunction and classical co-implication.

**Table 14.2** The understanding of structure connectives as logical operations

| | $\mathbf{I}$ | $\circ$ | $\bullet$ | $\odot$ | $\circledcirc$ | $\circledast$ | $*$ |
|---|---|---|---|---|---|---|---|
| in antecedent position | $\top$ | $\wedge$ | $\leftarrow_c$ | $\Diamond^-$ | $\blacklozenge^+$ | $\boxdiamond^-$ | $\neg$ |
| in succedent position | $\bot$ | $\rightarrow_c$ | $\vee$ | $\Box^+$ | $\blacksquare^-$ | $\boxbox^+$ | $\neg$ |

$$\sigma(X \vdash Y) = \sigma_1(X) \rightarrow \sigma_2(Y), \text{ where}$$

$$\begin{array}{rlrl}
\sigma_1(A) &= A & \sigma_2(A) &= A \\
\sigma_1(\mathbf{I}) &= \top & \sigma_2(\mathbf{I}) &= \bot \\
\sigma_1(X \circ Y) &= \sigma_1(X) \wedge \sigma_1(Y) & \sigma_2(X \circ Y) &= \sigma_1(X) \rightarrow_c \sigma_2(Y) \\
\sigma_1(X \bullet Y) &= \sigma_1(X) \leftarrow_c \sigma_2(Y) & \sigma_2(X \bullet Y) &= \sigma_2(X) \vee \sigma_2(Y) \\
\sigma_1(\odot X) &= \Diamond^- \sigma_1(X)) & \sigma_2(\odot X) &= \Box^+ \sigma_2(X) \\
\sigma_1(\circledcirc X) &= \blacklozenge^+ \sigma_1(X) & \sigma_2(\circledcirc X) &= \blacksquare^- \sigma_2(X) \\
\sigma_1(\circledast X) &= \boxdiamond^- \sigma_1(X) & \sigma_2(\circledast X) &= \boxbox^+ \sigma_2(X) \\
\sigma_1(*X) &= \neg \sigma_2(X) & \sigma_2(*X) &= \neg \sigma_1(X).
\end{array}$$

**Definition 14.4.4** The display sequent calculus $dBi\mathrm{N}$ is defined as the set consisting of the following axiom and rules:

Identity and cut

$$\frac{}{p \vdash p}\ (id) \qquad \frac{X \vdash A \quad A \vdash Y}{X \vdash Y}\ (cut)$$

Display rules

$$\frac{X \circ Y \vdash Z}{X \vdash Y \circ Z} \qquad \frac{X \circ Y \vdash Z}{Y \vdash X \circ Z} \qquad \frac{X \vdash Y \bullet Z}{X \bullet Y \vdash Z} \qquad \frac{X \vdash Y \bullet Z}{X \bullet Z \vdash Y}$$

$$\frac{\odot X \vdash Y}{X \vdash \odot Y} \qquad \frac{\circledcirc X \vdash Y}{X \vdash \circledcirc Y} \qquad \frac{\circledast X \vdash Y}{X \vdash \circledast Y} \qquad \frac{X \vdash Y}{*Y \vdash *X} \qquad \frac{X \vdash **Y}{X \vdash Y}$$

Structural rules

$$\frac{\mathbf{I} \circ X \vdash Y}{X \vdash Y} \qquad \frac{X \vdash Y \bullet \mathbf{I}}{X \vdash Y} \qquad \frac{X \vdash Y}{X \vdash Y \bullet Z} \qquad \frac{X \vdash Y}{X \circ Z \vdash Y}$$

$$\frac{X \vdash Y \bullet Z}{X \vdash Z \bullet Y} \qquad \frac{X \circ Z \vdash Y}{Z \circ X \vdash Y} \qquad \frac{X \vdash Y \bullet Y}{X \vdash Y} \qquad \frac{X \circ X \vdash Y}{X \vdash Y}$$

$$\frac{X \vdash (Y \bullet Z) \bullet W}{X \vdash Y \bullet (Z \bullet W)} \qquad \frac{(X \circ Y) \circ Z \vdash W}{X \circ (Y \circ Z) \vdash W}$$

$$\frac{A \vdash X}{\circledast A \vdash X} \quad \frac{X \vdash \circledast Y}{X \vdash Y} \quad \frac{X \vdash \circledast Y}{X \vdash \circledast \circledast Y}$$

Operational rules

$$\frac{X \vdash A \quad Y \vdash B}{X \circ Y \vdash A \wedge B} \quad \frac{A \circ B \vdash X}{A \wedge B \vdash X} \quad \frac{X \vdash A \bullet B}{X \vdash A \vee B} \quad \frac{A \vdash X \quad B \vdash Y}{A \vee B \vdash X \bullet Y}$$

$$\frac{\circledast X \circ A \vdash B}{X \vdash A \to B} \quad \frac{X \vdash A \quad B \vdash Y}{A \to B \vdash \circledast(X \circ Y)} \quad \frac{X \vdash B \quad A \vdash Y}{\circledast(X \bullet Y) \vdash B \leftarrow A} \quad \frac{B \vdash \circledast X \bullet A}{B \leftarrow A \vdash X}$$

$$\frac{*A \vdash X}{\triangleright A \vdash \odot X} \quad \frac{X \vdash \odot * A}{X \vdash \triangleright A} \quad \frac{*A \vdash X}{\triangleleft A \vdash * \odot * X} \quad \frac{X \vdash * \odot A}{X \vdash \triangleleft A}$$

$$\frac{\circledcirc * A \vdash Y}{\blacktriangleright A \vdash Y} \quad \frac{*Y \vdash *A}{\circledcirc * Y \vdash \blacktriangleright A} \quad \frac{* \circledcirc A \vdash Y}{\blacktriangleleft A \vdash Y} \quad \frac{Y \vdash *A}{* \circledcirc * Y \vdash \blacktriangleleft A}$$

Note that for any $\mathcal{L}$-formula $A$, the b-sequent $A \vdash A$ is derivable in $dBi\mathrm{N}$.

**Definition 14.4.5** (*Antecedent and succedent parts*) If $s = X \vdash Y$ is a b-sequent, then $X$ is an antecedent part (AP) of $s$ and $Y$ is a succedent part (SP) of $s$. Moreover:

- If $W \circ Z$ is an AP of $s$, then so are $W$ and $Z$;
- If $W \bullet Z$ is an SP of $s$, then so are $W$ and $Z$;
- If $W \circ Z$ is an SP of $s$, then $W$ is an AP and $Z$ is an SP of $s$;
- If $W \bullet Z$ is an AP of $s$, then $W$ is an AP and $Z$ is an SP of $s$;
- If $\odot W$ is an AP (SP) of $s$, then $W$ is an AP (SP) of $s$;
- If $\circledcirc W$ is an AP (SP) of $s$ then $W$ is an AP (SP) of $s$;
- If $\circledast W$ is an AP (SP) of $s$ then $W$ is an AP (SP) of $s$;
- If $*W$ is an AP (SP) of $s$, then $W$ is an SP (AP) of $s$.

**Definition 14.4.6** Two b-sequents are display equivalent (or structurally equivalent) if they are derivable from each other using only display rules of $dBi\mathrm{N}$.

**Theorem 14.4.7** (Display property) *Let $Z$ be a substructure of the b-sequent $s = X \vdash Y$. Then,*

- *If $Z$ is an AP of $s$, then there is a b-sequent $Z \vdash W$ that is display equivalent to $s$.*
- *If $Z$ is an SP of $s$, then there is a b-sequent $W \vdash Z$ that is display equivalent to $s$.*

**Theorem 14.4.8** (Cut elimination) *If a b-sequent $X \vdash Y$ is derivable in $dBi\mathrm{N}$, it is derivable without applications of (cut).*

***Proof*** Since $dBi\mathrm{N}$ is a proper display calculus, Belnap's general cut-elimination theorem applies. In particular, the principal cut-elimination steps, treated in Belnap's (1982) condition (C8), pose no problem. Here are, for example, the cases for two of the negative modalities, where the derivation on the left hand side is replaced by

the derivation on the right hand side, and where vertical dots indicate consecutive applications of display rules:

$$\dfrac{\dfrac{X \vdash \odot * A}{X \vdash \triangleright A} \quad \dfrac{*A \vdash Y}{\triangleright A \vdash \odot Y}}{X \vdash \odot Y}\,(cut) \qquad \dfrac{\dfrac{\dfrac{*A \vdash Y}{*Y \vdash **A}}{*Y \vdash A} \quad \dfrac{\dfrac{X \vdash \odot * A}{\odot X \vdash *A}}{\vdots}}{\dfrac{*Y \vdash * \odot X}{\odot X \vdash Y}}\,(cut)}{X \vdash \odot Y}$$

$$\dfrac{\dfrac{X \vdash *A}{* \circledcirc *X \vdash \blacktriangleleft A} \quad \dfrac{* \circledcirc A \vdash Y}{\blacktriangleleft A \vdash Y}}{* \circledcirc *X \vdash Y}\,(cut) \qquad \dfrac{\dfrac{\dfrac{* \circledcirc A \vdash Y}{*Y \vdash \circledcirc **A}}{*Y \vdash \circledcirc A} \quad \dfrac{X \vdash *A}{\vdots}}{\dfrac{\circledcirc * Y \vdash A}{\circledcirc * Y \vdash *X}\,(cut)}}{\dfrac{*Y \vdash \circledcirc * X}{* \circledcirc *X \vdash ** Y}}$$
$$* \circledcirc *X \vdash Y$$

**Definition 14.4.9** A b-sequent $X \vdash Y$ is valid with respect to a class of $BiN$-frames iff the $\mathcal{L}^{\neg}$-formula $\sigma(X \vdash Y)$ is valid with respect to that class.

Note that there are b-sequents derivable in $dBiN$ that are not valid with respect to the class of all $BiN$-frames under the translation $\sigma'$ that is like $\sigma$ except that $\rightarrow_c$ is replaced by $\rightarrow$ and $\leftarrow_c$ is replaced by $\leftarrow$. Consider, for example, the following derivations[7]:

$$\dfrac{\dfrac{\dfrac{p \vdash p}{*p \vdash *p}}{*p \circ q \vdash *p}}{*p \vdash q \circ *p} \qquad \dfrac{\dfrac{\dfrac{p \vdash p}{*p \vdash *p}}{*p \vdash *p \bullet q}}{*p \bullet q \vdash *p}$$

The $\mathcal{L}^{\neg}$-formulas $\sigma'(*p \vdash q \circ *p)$ $(= \neg p \rightarrow (q \rightarrow \neg p))$ and $\sigma'(*p \bullet q \vdash *p)$ $(= (\neg p \leftarrow q) \rightarrow \neg p)$ fail to be valid because heredity does not hold for classically negated formulas.

**Theorem 14.4.10** (Soundness) *If a b-sequent $X \vdash Y$ is derivable in $dBiN$, then $X \vdash Y$ is valid with respect to the class of all $BiN$ -frames.*

***Proof*** By induction on derivations in $dBiN$. The claim obviously holds for instances of $(id)$, and it is not difficult to see that the other sequent rules preserve validity. We present here the cases of the right and left introduction rules for $\rightarrow$ and $\leftarrow$, the display rules for $\circledcirc$, and some rules for the modal operators. The structural rules for

[7] I am grateful to Sergey Drobyshevich for pointing out to me the problem with $\sigma'$.

$\circledast$ are validity preserving because of the persistence property and the reflexivity and transitivity of the relation $\leq$ in $Bi$N-models.

Consider the rule

$$\frac{\circledast X \circ A \vdash B}{X \vdash A \rightarrow B}$$

and assume that $(\star)\ \sigma(\circledast X \circ A \vdash B)$, i.e., $(\Diamond^{-}\sigma_1(X) \wedge A) \rightarrow B$, is valid, but $\sigma(X \vdash (A \rightarrow B))$, i.e., $\sigma_1(X) \rightarrow (A \rightarrow B)$, is not. Then there is a model $(W, \leq, \frown, \smile, \models)$ and there are $x, y \in W$ with $x \leq y$, $x \not\models \sigma_1(X) \rightarrow (A \rightarrow B)$, $y \models \sigma_1(X)$, and $y \not\models (A \rightarrow B)$. Hence there is a $z \in W$ with $y \leq z$, $z \models A$, and $z \not\models B$. By $(\star)$, $z \models B$, and we obtain a contradiction.

The case of the rule

$$\frac{B \vdash \circledast X \bullet A}{B \leftarrow A \vdash X}$$

is similar. Assume that $(\star)\ B \rightarrow (\Box^{+}\sigma_2(X) \vee A)$ is valid, but $(B \leftarrow A) \rightarrow \sigma_2(X)$ is not. Then there exists a model $(W, \leq, \frown, \smile, \models)$ and there are $x, y \in W$ with $x \leq y$, $(\star\star)\ y \models B \leftarrow A$, and $y \not\models \sigma_2(X)$. By $(\star\star)$, there is a $z \in W$ with $z \leq y$, $z \models B$, and $z \not\models A$. By $(\star)$ , $z \models \Box^{+}\sigma_2(X)$ and thus $y \models \sigma_2(X)$, and we obtain a contradiction.

The case of the rule

$$\frac{X \vdash A \quad B \vdash Y}{(A \rightarrow B) \vdash \circledast(X \circ Y)}$$

basically amounts to seeing that $A \rightarrow B$ is a strict Boolean implication. Assume that $(\star)\ \sigma_1(X) \rightarrow A$ and $B \rightarrow \sigma_2(Y)$ are valid, but $(A \rightarrow B) \rightarrow \Box^{+}(\sigma_1(X) \rightarrow_c \sigma_2(Y))$ is not valid. Then there is a model $(W, \leq, \frown, \smile, \models)$ and there exist $x, y \in W$ with $x \leq y$, $(\star\star)\ y \models (A \rightarrow B)$, and $y \not\models \Box^{+}(\sigma_1(X) \rightarrow_c \sigma_2(Y))$. Hence, there is a $z \in W$ with $y \leq z$ and $z \models \sigma_1(X)$ but $z \not\models \sigma_2(Y)$. By $(\star)$ and $(\star\star)$, $z \models \sigma_2(Y)$, and we obtain a contradiction.

The case of the rule

$$\frac{X \vdash B \quad A \vdash Y}{\circledast(X \bullet Y) \vdash (B \leftarrow A)}$$

is similar. Assume that $(\star)\ \sigma_1(X) \rightarrow B$ and $A \rightarrow \sigma_2(Y)$ are valid, but that the formula $\Diamond^{-}(\sigma_1(X) \leftarrow_c \sigma_2(Y)) \rightarrow (B \leftarrow A)$ is not valid, and there exists a model break $(W, \leq, \frown, \smile, \models)$, there are $x, y \in W$ with $x \leq y$, and $(\star\star)\ y \models \Diamond^{-}(\sigma_1(X) \leftarrow_c \sigma_2(Y))$ so that there is a $z \in W$ with $z \leq y$, $z \models \sigma_1(X)$, and $z \not\models \sigma_2(Y)$. Moreover, $y \not\models (B \leftarrow A)$. Thus for every $w \in W$, if $w \leq y$, then (i) $w \not\models B$ or (ii) $w \models A$. By $(\star)$, $z \models B$ and $z \not\models A$, which gives us a contradiction.

Now we consider the rule

$$\frac{*A \vdash X}{\rhd A \vdash \odot X}$$

and assume that $(\star)\ \neg A \rightarrow \sigma_2(X)$ is valid whereas $\rhd A \rightarrow \Box^{+}\sigma_2(X)$ is not valid. Then there is a model $(W, \leq, \frown, \smile, \models)$ and there are $x, y \in W$ with $x \leq y$, $(\star\star)$

$y \models \rhd A$, and $y \not\models \Box^+\sigma_2(X)$. Hence, there is a $z \in W$ with $y \frown z$ and $z \not\models \sigma_2(X)$. By $(\star\star)$, $z \not\models A$ and hence, by $(\star)$, $z \models \sigma_2(X)$, so that we get a contradiction.

Now we consider the rule

$$\frac{*A \vdash X}{\rhd A \vdash \odot X}$$

and assume that $(\star)$ $\neg A \to \sigma_2(X)$ is valid whereas $\rhd A \to \Box^+\sigma_2(X)$ is not valid. Then there is a model $(W, \leq, \frown, \smile, \models)$ and there are $x, y \in W$ with $x \leq y$, $(\star\star)$ $y \models \rhd A$, and $y \not\models \Box^+\sigma_2(X)$. Hence, there is a $z \in W$ with $y \frown z$ and $z \not\models \sigma_2(X)$. By $(\star\star)$, $z \not\models A$ and hence, by $(\star)$, $z \models \sigma_2(X)$, so that we get a contradiction.

As a final example, consider the rule

$$\frac{*Y \vdash *A}{\circledcirc * Y \vdash \blacktriangleright A}$$

and assume $(\star)$ $\neg\sigma_2(Y) \to \neg A$, is valid, but $\blacklozenge^+\neg\sigma_2(Y) \to \blacktriangleright A$, is not. Then there exists a model $(W, \leq, \frown, \smile, \models)$ and there exist $x, y \in W$ with $x \leq y$, $y \models \blacklozenge^+\neg\sigma_2(Y)$ but $(\star\star)$ $y \not\models \blacktriangleright A$. Hence there is a $z \in W$ with $y \smile z$ and $z \models \neg\sigma_2(Y)$. By $(\star)$, $z \not\models A$, but by $(\star\star)$, for every $t \in W$, $y \smile t$ implies $t \models A$.

**Theorem 14.4.11** *If an o-sequent $X \vdash Y$ is derivable in $\delta Bi$*N*, then the b-sequent $\tau_1(X) \vdash \tau_2(Y)$ is derivable in $dBi$*N.

***Proof*** By induction on derivations in $\delta Bi$N. Note that the derivation of $\tau_1(X) \vdash \tau_2(Y)$ in $dBi$N need not be cut-free. Consider, for example the display rule

$$\frac{X > Y \vdash Z}{X \vdash Z; Y}$$

from $\delta Bi$N. We have the following derivation in $dBi$N:

$$\dfrac{\dfrac{\dfrac{\dfrac{\dfrac{\dfrac{\dfrac{\tau_1(X) \vdash \tau_1(X) \quad \tau_2(Y) \vdash \tau_2(Y)}{\circledast(\tau_1(X) \bullet \tau_2(Y)) \vdash \tau_1(X) \leftarrow \tau_2(Y)} \quad \tau_1(X) \leftarrow \tau_2(Y) \vdash \tau_2(Z)}{\circledast(\tau_1(X) \bullet \tau_2(Z)) \vdash \tau_2(Y)}\,(cut)}{\tau_1(X) \bullet \tau_2(Z) \vdash \circledast\tau_2(Y)}}{\tau_1(X) \bullet \tau_2(Z) \vdash \tau_2(Y)}}{\tau_1(X) \vdash \tau_2(Z) \bullet \tau_2(Y)}}{\tau_1(X) \vdash \tau_2(Z) \vee \tau_2(Y)}$$

We present two other cases. Consider the rule

$$\frac{X \vdash \sharp Y}{Y \vdash \flat X}$$

Then we have the following derivation in $dBi$N:

$$\dfrac{\dfrac{\dfrac{\dfrac{\tau_1(X) \vdash \triangleright\tau_1(Y) \quad \dfrac{\dfrac{\tau_1(Y) \vdash \tau_1(Y)}{*\tau_1(Y) \vdash *\tau_1(Y)}}{\triangleright\tau_1(Y) \vdash \odot * \tau_1(Y)}}{\tau_1(X) \vdash \odot * \tau_1(Y)}}{\odot\tau_1(X) \vdash *\tau_1(Y)}}{\vdots}}{\dfrac{\dfrac{**\odot\tau_1(X) \vdash *\tau_1(Y)}{\tau_1(Y) \vdash * \odot \tau_1(X)}}{\tau_1(Y) \vdash \triangleleft\tau_1(X)}} \; (cut)$$

For the rule

$$\dfrac{A \vdash Y}{\flat Y \vdash \blacktriangleleft A}$$

we have the following derivation in $dBi$N:

$$\dfrac{\dfrac{\dfrac{A \vdash \tau_2(Y)}{*\tau_2(Y) \vdash *A}}{* \circledcirc ** \tau_2(Y) \vdash \blacktriangleleft A}}{\dfrac{\vdots}{* \circledcirc \tau_2(Y) \vdash \blacktriangleleft A}}{\blacktriangleleft \tau_2(Y) \vdash \blacktriangleleft A}$$

□

**Theorem 14.4.12** *An o-sequent $A \vdash B$ is derivable in $\delta Bi$*N *iff the b-sequent $A \vdash B$ is derivable in $dBi$*N.

***Proof*** Direction from right to left: If $A \vdash B$ is derivable in $dBi$N, then, by Theorem 14.4.10, $A \vdash B$ is valid with respect to the class of all $Bi$N-frames, and by Theorem 14.2.11, $A \vdash B$ is derivable in $\delta Bi$N. Direction from left to right: By Theorem 14.4.11. □

**Corollary 14.4.13** *A b-sequent $A \vdash B$ is derivable in $dBiN$ if and only if $A \vdash B$ is valid with respect to the class of all $BiN$-frames.*

The o-sequents listed in Table 14.3 are valid on a $Bi$N-frame if and only if the first-order conditions given with them are satisfied. In each case a corresponding structural sequent rule is presented in linear notation. Cut-elimination holds for extensions of $dBi$N with any combination of these rules because the resulting systems are proper display calculi.

We now define a notion of frame correspondence that is weaker than the standard notion of frame correspondence, so that we can state structural rules that weakly correspond with double negation elimination for negation as impossibility.

**Table 14.3** Some structural rules over $dBi\mathrm{N}$

| | |
|---|---|
| DNI $A \vdash \rhd \rhd A$ | $\forall x \forall y (x \frown y \Rightarrow y \frown x)$ |
| (DNI) $X \vdash \odot * Y / X \vdash * \odot Y$ | |
| Con $(A \to B) \vdash (\rhd B \to \rhd A)$ | $\forall x \forall y (x \frown y \Rightarrow$ |
| | $\exists z (x \leq z \& y \leq z \& x \frown z))$ |
| (Con) $X \vdash \odot * (Y \circ Z) / X \vdash Y \circ \odot * Z$ | |
| LNC $(A \wedge \rhd A) \vdash B$ | $\forall x \, x \frown x$ |
| (LNC) $X \vdash \odot * Y / X \vdash Y$ | |
| $\rhd \top \bot \rhd \top \vdash \bot$ | $\forall x \exists y \, x \frown y$ |
| $(\rhd \top \bot) X \vdash \odot * \mathbf{I} / X \vdash \mathbf{I}$ | |
| LEM $B \vdash A \vee \rhd A$ | $\forall x \forall y (x \frown y \Rightarrow y \leq x)$ |
| (LEM) $X \circ Y \vdash Z / Y \vdash \odot * X \bullet Z$ | |
| FDM $\rhd (A \wedge B) \vdash (\rhd A \vee \rhd B)$ | $\forall x \forall y \forall z ((x \frown y \& x \frown z) \Rightarrow$ |
| | $\exists w (y \leq w \& z \leq w \& x \frown w))$ |
| (FDM) $X \vdash \odot * (Y \bullet Z) / X \vdash \odot * Y \bullet \odot * Z)$ | |

**Definition 14.4.14** Let $[E]$ be a property of $Bi\mathrm{N}$-frames, and let $\mathcal{E}$ be the class of all $Bi\mathrm{N}$-frames with property $[E]$. A sequent $s$ (sequent rule $r$) weakly corresponds with $[E]$ if the following holds: 1. For every $\mathcal{F} \in \mathcal{E}$, $A$ is valid (preserves validity) on $\mathcal{F}$. 2. For every $Bi\mathrm{N}$-frame $\mathcal{F} \notin \mathcal{E}$, there is an $\mathcal{F}' \in \mathcal{E}$ such that $A$ in not valid (does not preserve validity) on $\mathcal{F}'$.

The double negation sequent DNE weakly corresponds with the frame property

$$[\mathrm{DNE}'] \quad \forall x \exists y (x \frown y \& \forall z (y \frown z \Rightarrow z = x)).$$

If a frame $\mathcal{F}$ belongs to $\mathcal{DNE}'$, it is easy to see that DNE is valid on $\mathcal{F}$. If $\mathcal{F} = (W, \leq, \frown, \smile) \notin \mathcal{DNE}'$, then $\exists x \forall y (x \frown y \Rightarrow \exists z (y \frown z \& z \neq x))$. We define the structure $\mathcal{F}' = (W, \leq', \frown, \smile)$ by setting $\leq' := \{(x, x) \mid x \in W\}$. Clearly $\leq'$ is a preorder on $W$, and the conditions (i) and (ii) from Definition 14.2.1 are satisfied, so that $\mathcal{F}'$ is a $Bi\mathrm{N}$-frame that does not belong to $\mathcal{DNE}'$. Given the definition of $\leq'$, $\exists x \forall y (x \frown y \Rightarrow \exists z (y \frown z \& z \not\leq' x))$. Hence $\mathcal{F}' \notin \mathcal{DNE}$ and thus DNE is not valid on $\mathcal{F}'$ (because DNE corresponds with [DNE]).

We may note that in classical first-order logic with identity, $[\mathrm{DNE}']$ is equivalent with the conjunction of two conditions [DNE1] and [DNE2]. Condition [DNE1] weakly corresponds with a structural sequent rule (DNE1), and [DNE2] corresponds with a structural sequent rule (DNE2):

| | |
|---|---|
| [DNE1] $\forall x \forall y \forall z ((y \frown x \& z \frown y) \Rightarrow x = z)$ | (DNE1) $X \vdash Y / \odot \odot X \vdash Y$ |
| [DNE2] $\forall x \exists y \, x \frown y$ | (DNE2) $X \vdash \odot * \odot * Y / X \vdash Y$ |

Suppose that $[\mathrm{DNE}']$ is satisfied. Then clearly DNE2 is also satisfied. Choose $a, b, c$ such that $a \frown b$ and $b \frown c$. It must be shown that $c = a$. By $[\mathrm{DNE}']$, there must be a

$b_0$ such that $a \frown b_0$, and (i) $\forall z(b_0 \frown z) \Rightarrow z = a)$. By (i), $a$ is the only $\frown$-successor of $b_0$. By [DNE′], (ii) there exists an $a_0$ for the given $b_0$ with $b_0 \frown a_0$. Moreover, (iii) $\forall z((a_0 \frown z) \Rightarrow z = b_0)$. By (i) and (ii), we have (iv) $a_0 = a$. By (iii) and (iv), $b_0$ is the only $\frown$-successor of $a$. Since $a \frown b$, it follows that $b = b_0$. But then also, by $b \frown c$, it follows $b_0 \frown c$. Therefore, by (i), $c = a$. For the other direction, assume that [DNE1] and [DNE2] are satisfied. Then, by [DNE2], for a given state $a$, we can choose a $b$ with $a \frown b$. If for any given $c$ we have $b \frown c$, then $a = c$ by [DNE1].

All structural sequent rules listed in Table 14.3 preserve validity with respect to the class of $Bi$N-frames satisfying the corresponding frame condition. The combination of (DNE1) and (DNE2) weakly corresponds with [DNE′]. While $dBi$N plus (DNE1) and (DNE2) is sound with respect to $\mathcal{DNE}'$, the question of completeness is open. We here present derivations of all sequents listed sequents in Table 14.3 and a derivation of $\rhd \rhd A \vdash A$ in $dBi$N plus (DNE1) and (DNE2):

$$\dfrac{\dfrac{\dfrac{\dfrac{\dfrac{\dfrac{\dfrac{\dfrac{A \vdash A}{*A \vdash *A}}{\rhd A \vdash \odot * A}}{\rhd A \vdash * \odot A}}{* * \odot A \vdash * \rhd A}}{* * \odot * A \vdash * * * \rhd A}}{* * \rhd A \vdash * \odot A}}{\odot A \vdash * \rhd A}}{A \vdash \odot * \rhd A}}{A \vdash \rhd \rhd A} \text{ (DNI)} \qquad \dfrac{\dfrac{\dfrac{\dfrac{\dfrac{\dfrac{\dfrac{A \vdash A \quad B \vdash B}{(A \to B) \vdash \circledast(A \circ B)}}{(A \to B) \vdash (A \circ B)}}{((A \to B) \circ A) \vdash B}}{*B \vdash *((A \to B) \circ A)}}{\rhd B \vdash \odot * ((A \to B) \circ A)}}{\rhd B \vdash ((A \to B) \circ \odot * A)}}{(\rhd B \circ (A \to B)) \vdash \odot * A}}{(\rhd B \circ (A \to B)) \vdash \rhd A}}{(A \to B) \vdash (\rhd B \circ \rhd A)} \text{ (Con)}$$

$$\dfrac{\dfrac{\dfrac{\dfrac{A \vdash A}{(A \circ B) \vdash A}}{\rhd A \vdash \odot * (A \circ B)}}{\rhd A \vdash (A \circ B)}}{(A \circ \rhd A) \vdash B}}{(A \wedge \rhd A) \vdash B} \text{ (LNC)} \qquad \dfrac{\dfrac{\dfrac{\dfrac{\dfrac{\dfrac{p \vdash p}{\circledast \mathbf{I} \circ p \vdash p}}{\mathbf{I} \vdash (p \to p)}}{*(p \to p) \vdash * \mathbf{I}}}{\rhd(p \to p) \vdash \odot * \mathbf{I}}}{\rhd(p \to p) \vdash \mathbf{I}}}{\rhd(p \to p) \vdash \mathbf{I} \bullet (p \leftarrow p)}}{\rhd(p \to p) \vdash (p \leftarrow p)} (\rhd \top \bot)$$

$$\dfrac{\dfrac{\dfrac{\dfrac{\dfrac{\dfrac{\dfrac{A \vdash A}{A \circ B \vdash A}}{B \vdash \odot * A \bullet A}}{B \bullet A \vdash \odot * A}}{B \bullet A \vdash \rhd A}}{B \vdash A \bullet \rhd A}}{B \vdash (A \vee \rhd A)} \text{(LEM)} \qquad \dfrac{\dfrac{\dfrac{\dfrac{\dfrac{\dfrac{\dfrac{\dfrac{\dfrac{A \vdash A}{A \vdash (A \bullet B)}}{(A \circ B) \vdash (A \bullet B)}}{(A \wedge B) \vdash (A \bullet B)}}{\rhd(A \wedge B) \vdash \odot * (A \bullet B)}}{\rhd(A \wedge B) \vdash \odot * A \bullet \odot * B}}{\rhd(A \wedge B) \bullet \odot * B \vdash \odot * A}}{\rhd(A \wedge B) \bullet \odot * B \vdash \rhd A}}{\rhd(A \wedge B) \vdash \odot * B \bullet \rhd A}}{\rhd(A \wedge B) \bullet \rhd A \vdash \odot * B}}{\rhd(A \wedge B) \bullet \rhd A \vdash \rhd B}}{\rhd(A \wedge B) \vdash \rhd A \bullet \rhd B}}{\rhd(A \wedge B) \vdash (\rhd A \vee \rhd B)} \text{(FDM)}$$

$$\dfrac{\dfrac{\dfrac{\dfrac{\dfrac{\dfrac{A \vdash A}{* A \vdash * A}}{\odot \odot * A \vdash * A}}{\odot * A \vdash \odot * A}}{\odot * A \vdash \rhd A}}{* \rhd A \vdash * \odot * A}}{\rhd \rhd A \vdash \odot * \odot * A} \text{(DNE1)} \qquad \text{(DNE2)}$$

$$\rhd \rhd A \vdash A$$

## 14.5 Conclusion

This paper is not intended to pit Onishi's solution to his two problems concerning substructural negations against the approach to his first problem presented here. Both strategies have their advantages and disadvantages. Onishi's solution dispenses with a structure connective that represents Boolean negation in the language of structures. (A general philosophical discussion about Boolean negation and its relationship to non-Boolean negation connectives is beyond the scope of this paper.) Moreover, in applications where one is justified in identifying negation as impossibility and negation as unnecessity, Onishi's solution works well. For such applications of functional modalities one may consult, e.g., Segerberg (1986) and Kravtsov (2002). The present paper shows that a different language of structures allows one to state structural sequent rules (DNE1) and (DNE2) such that there is a frame property, [DNE′], to which both double negation elimination for negation as impossibility and the pair consisting of (DNE1) and (DNE2) weakly correspond. The result of adding the structural rules (DNE1) and (DNE2) to the sequent calculus $dBi$N is sound with respect to the class of frames satisfying [DNE′], the question of completeness is open. The structural language of sequents used in $dBi$N lays down the meaning of its structure connectives by means of separated structural rules.[8]

**Acknowledgements** Earlier versions of this paper have been presented at the conference *The Logic of Paradox, after 40 years*, Bochum, November 2019 and at the ILLC, Amsterdam, February 2020. I would like to thank the audiences for critical and useful remarks, in particular, Takuro Onishi,

[8] In Sergey Drobyshevich, Sergei Odintsov, and Heinrich Wansing, "Moisil's modal logic and related systems", in: Katalin Bimbó (Ed.), *Relevance Logics and other Tools for Reasoning. Essays in Honor of J. Michael Dunn*, College Publications, London, 2022, 150–177, it is observed that Grigore Moisil already in his essay "Logique modale", *Disquisitiones Mathematicae et Physicae* 2 (1942), 3–98, introduced a propositional logic equivalent to Heyting-Brouwer logic.

Hitoshi Omori, Sergei Odintsov, Sergey Drobyshevich, João Marcos, Graham Priest, and Franz Berto. Special thanks go to Grigory Olkhovikov and Sergey Drobyshevich for their advice and to two anonymous referees for their very helpful comments.

## Appendix

**Definition 14.5.1** Let $\Delta, \Gamma$ be two sets of $\mathcal{L}$-formulas, i.e., $\neg$-free formulas. We write $\Delta \vdash \Gamma$ if the sequent $(\ldots(A_1 \wedge \ldots) \wedge A_n) \vdash (\ldots(B_1 \vee \ldots) \vee B_m)$ is derivable in $\delta Bi\mathrm{N}$ for some $A_1, \ldots, A_n \in \Delta$ and $B_1, \ldots, B_m \in \Gamma$, and we write $\Delta \nvdash \Gamma$ if that is not the case. A pair $(\Delta, \Gamma)$ is *maximal consistent* if $\Delta \cup \Gamma = Form_{\mathcal{L}}$ and $\Delta \nvdash \Gamma$.

**Lemma 14.5.2** (Onishi, 2015) *Let $\Delta, \Gamma$ be two sets of $\mathcal{L}$-formulas and $\Delta \nvdash \Gamma$. Then there is a maximal consistent pair $(\Delta', \Gamma')$ with $\Delta \subseteq \Delta'$ and $\Gamma \subseteq \Gamma'$.*

**Definition 14.5.3** (Onishi, 2015; Drobyshevich, 2018) The canonical model

$$\mathcal{M}^* = (W^*, \leq^*, \frown^*, \smile^*, \models^*)$$

is defined as follows:

- $W^* = \{\Delta \mid (\Delta, Form(_{\mathcal{L}}) \setminus \Delta)$ is a maximal consistent pair$\}$;
- $\Delta \leq^* \Gamma$ iff $\Delta \subseteq \Gamma$;
- $\Delta \frown^* \Gamma$ iff $\forall A(\triangleright A \in \Delta \Rightarrow A \notin \Gamma)$;
- $\Delta \smile^* \Gamma$ iff $\exists A(\blacktriangleright A \in \Delta$ and $A \notin \Gamma)$;
- $\Delta \models^* p$ iff $p \in \Delta$.

The canonical model is a $Bi\mathrm{N}$-model.

**Lemma 14.5.4** *For any $\Delta \in W^*$ in $\mathcal{M}^*$ and any $\mathcal{L}^*$-formula $A$, $\Delta \models A$ iff $A \in \Delta$.*

Completeness follows from these two lemmas in the usual way.

**Theorem 14.5.5** (Completeness) *If an o-sequent is valid with respect to the class of all $Bi\mathrm{N}$-models, then it is derivable in $\delta Bi\mathrm{N}$.*

## References

Almukdad, A., & Nelson, D. (1984). Constructible falsity and inexact predicates. *Journal of Symbolic Logic, 49*, 231–233.

Anderson, A. R., Belnap, N. D., & Dunn, J. M. (1992). *Entailment: The logic of relevance and necessity* (Vol. 2). Princeton: Princeton University Press.

Belnap, N. D. (1982). Display logic. *Journal of Philosophical Logic, 11*, 375–417. Reprinted with minor changes as §62 of Anderson, Belnap, and Dunn (1992).

Berto, F. (2014). A modality called 'negation'. *Mind, 124*, 761–793.
Berto, F., & Restall, G. (2019). Negation on the Australian plan. *Journal of Philosophical Logic, 48*, 1119–1144.
Birkhoff, G., & von Neumann, J. (1936). The logic of quantum mechanics. *The Annals of Mathematics, 37*, 823–843.
Ciabattoni, A., & Ramanayake, R. (2016). Power and limits of structural display rules. *ACM Transactions on Computational Logic, 17*, 1–39.
Ciabattoni, A., Ramanayake, R., & Wansing, H. (2014). Hypersequent and display calculi-a unified perspective. *Studia Logica, 102*, 1245–1294.
De, M., & Omori, H. (2018). There is more to negation than modality. *Journal of Philosophical Logic, 47*, 281–299.
Došen, K. (1984). Negative modal operators in intuitionistic logic. *Publications de l'Institut Mathématique. Nouvelle Série, 35*(49), 3–14.
Došen, K. (1986). Negation as a modal operator. *Reports on Mathematical Logic, 20*, 15–27.
Došen, K. (1988). Sequent systems and groupoid models I. *Studia Logica, 47*, 353–389.
Došen, K. (1999). Negation in the light of modal logic. In D. Gabbay & H. Wansing (Eds.), *What is negation?* (pp. 77–86). Dordrecht: Kluwer.
Drobyshevich, S. (2018). On displaying negative modalities. *Logic and Logical Philosophy, 27*, 161–192.
Dunn, J. M. (1993). Star and perp: Two treatments of negation. *Philosophical Perspectives (Philosophy of Language and Logic), 7*, 331–357.
Dunn, J. M. (1996). Generalized ortho negation. In H. Wansing (Ed.), *Negation. A notion in focus* (pp. 3–26). Berlin: De Gruyter.
Dunn, J. M. (1999). A comparative study of various model-theoretic treatments of negation: A history of formal negation. In D. Gabbay & H. Wansing (Eds.), *What is negation?* (pp. 23–51). Dordrecht: Kluwer.
Dunn, J. M., & Zhou, C. (2005). Negation in the context of gaggle theory. *Studia Logica, 80*, 235–264.
Goldblatt, R. (1974). Semantic analysis of orthologic. *Journal of Philosophical Logic, 3*, 19–35.
Goré, R. (1995, January 1). *Intuitionistic logic redisplayed*. Tech. Rep. Canberra: Automated Reasoning Project, Australian National University.
Goré, R. (2000). Dual intuitionistic logic revisited. In R. Dyckhoff (Ed.), *Proceedings of the international conference on automated reasoning with analytic tableaux and related methods* (pp. 252–267). Berlin: Springer.
Goré, R., & Shillito, I. (2020). Bi-intuitionistic logics: A new instance of an old problem. In N. Olivetti, R. Verbrugge, S. Negri, & G. Sandu (Eds.), *Advances in modal logic* (Vol. 13, pp. 269–288). London: College Publications.
Horn, L., & Wansing, H. (2015). *Negation.* [2022] *The Stanford Encyclopedia of Philosophy* (Winter 2022 Edition), Edward N. Zalta & Uri Nodelman (eds.). https://plato.stanford.edu/archives/win2022/entries/negation
Kamide, N., & Wansing, H. (2015). *Proof theory for N4-related paraconsistent logics*. London: College Publications.
Kravtsov, A. (2002). Polymodal logics of commuting functions. *Logic Journal of the IGPL, 10*, 517–533.
Kripke, S. A. (1959). A completeness theorem in modal logic. *Journal of Symbolic Logic, 24*, 1–14.
Kripke, S. A. (1963a). Semantical analysis of intuitionistic logic I. In J. Crossley & M. Dummett (Eds.), *Formal systems and recursive functions (proceedings of the eighth logic colloquium at Oxford, July 1963)* (pp. 92–129). Amsterdam: North Holland Publishing Company.
Kripke, S. A. (1963b). Semantical analysis of modal logic I. Normal modal propositional calculi. *Zeitschrift für mathematische Logik und Grundlagen der Mathematik, 9*, 67–96.
Kripke, S. A. (1963c). Semantical considerations on modal logic. *Acta Philosophica Fennica, 16*, 83–94.

Kripke, S. A. (1965). Semantical analysis of modal logic II. Non-normal modal propositional calculi. In J. W. Addison, L. Henkin & A. Tarski (Eds.), *The theory of models (proceedings of the 1963 international symposium at Berkeley)* (pp. 206–220). Amsterdam: North Holland Publishing Company.

Lahav, O., Marcos, J., & Zohar, Y. (2017). Sequent systems for negative modalities. *Logica Universalis, 11*, 345–382.

Łukowski, P. (1996). Modal interpretation of Heyting-Brouwer logic. *Bulletin of the Section of Logic, 25*, 80–83.

Mares, E. (2004). *Relevant logic: A philosophical interpretation*. Cambridge: Cambridge University Press.

Odintsov, S., & Wansing, H. (2021). Routley star and hyperintensionality. *Journal of Philosophical Logic, 50*, 33–56.

Omori, H., & Wansing, H. (2020). An extension of connexive logic C. In N. Olivetti, R. Verbrugge, S. Negri, & G. Sandu (Eds.), *Advances in modal logic* (Vol. 13, pp. 503–522). London: College Publications.

Onishi, T. (2015). Substructural negations. *Australasian Journal of Logic, 12*(4), 1–27.

Parisi, A. (2022). A hypersequent solution to the inferentialist problem of modality. *Erkenntnis, 87*, 1605–1633.

Poggiolesi, F. (2010). *Gentzen calculi for modal propositional logic*. Springer.

Priest, G. (2008). *An introduction to non-classical logic: From if to is* (2nd ed.). Cambridge: Cambridge University Press.

Restall, G. (1999). Negation in relevant logics (How I stopped worrying and learned to love the Routley star). In D. Gabbay & H. Wansing (Eds.), *What is negation?* (pp. 53–76). Dordrecht: Kluwer.

Segerberg, K. (1986). Logics with functional alternative relations. *Notre Dame Journal of Formal Logic, 27*, 504–522.

Vakarelov, D. (1977). *Theory of negation in certain logical systems: Algebraic and semantic approach*. Unpublished doctoral dissertation. University of Warsaw.

Vakarelov, D. (1989). Intuitive semantics for some three-valued logics connected with information, contrariety and subcontrariety. *Studia Logica, 48*, 565–575.

Vakarelov, D. (1989). Consistency, completeness and negations. In G. Priest, R. Routley, & J. Norman (Eds.), *Paraconsistent logic: Essays on the inconsistent* (pp. 328–368). Munich: Philosophia Verlag.

Wansing, H. (1994). Sequent calculi for normal modal propositional logics. *Journal of Logic and Computation, 4*, 125–142.

Wansing, H. (1998). *Displaying modal logic*. Dordrecht: Kluwer Academic Publishers.

Wansing, H. (2005). Connexive modal logic. In R. Schmidt, I. Pratt-Hartmann, M. Reynolds, & H. Wansing (Eds.), *Advances in modal logic* (Vol. 5, pp. 367–383). London: King's College Publications.

Wansing, H. (2006). *Connexive logic*. [2023] *The Stanford Encyclopedia of Philosophy* (Summer 2023 Edition), Edward N. Zalta & Uri Nodelman (eds.). https://plato.stanford.edu/archives/sum2023/entries/logic-connexive

Wansing, H. (2008). Constructive negation, implication, and co-implication. *Journal of Non-Classical Logics, 18*, 341–364.

Wansing, H. (2016). Falsification, natural deduction, and bi-intuitionistic logic. *Journal of Logic and Computation, 26*, 425–450.

Wolter, F. (1998). On logics with coimplication. *Journal of Philosophical Logic, 27*, 353–387.

# Chapter 15
# New(ish) Foundations for Theories of Entailment

**Yale Weiss**

**Abstract** In this chapter, I present a systematic study of theories of entailment as developed in relevant modal logics of varying strength. I consider several relevant normal modal logics and the properties of entailment, as defined as necessitated relevant implication, in these systems. Fine-Urquhart style semantics are provided for these relevant modal logics and proofs of soundness and completeness are given. The semantics is used to prove admissibility results for certain weak relevant modal theories. Connections between some of these relevant modal logics and preexisting theories of entailment are pointed out.

**Keywords** Admissibility · Entailment · Modal logic · Operational semantics · Relevant modal logic

## 15.1 Introduction

The story of 20th century modal logic is widely known, if mainly in broad brush-strokes. It goes something like this. Lewis and Langford (1932/1959) presented a number of axiomatic systems (including perennial favorites like **S4** and **S5**) for strict implication, i.e., necessitated implication. The strict implication formulations were cumbersome and unilluminating, but, through the work of logicians including Gödel (1933/1986a) and Lemmon (1957), they were eventually replaced by axiomatizations which separated out the specifically modal and non-modal elements in the way that is now familiar. In the first half of the 20th century, syntax dominated. But in the second

Y. Weiss (✉)
The Saul Kripke Center, The Graduate Center, CUNY, 365 5th Avenue, Room 7118, New York, NY 10016, USA
e-mail: yweiss@gradcenter.cuny.edu

© The Author(s), under exclusive license to Springer Nature Switzerland AG 2024

Y. Weiss and R. Birman (eds.), *Saul Kripke on Modal Logic*, Outstanding Contributions to Logic 30, https://doi.org/10.1007/978-3-031-57635-5_15

half of the 20th century, through the work of Kripke (1959, 1963, 1965) especially (though not exclusively), modal logic obtained a robust and intuitive semantics and model theory.[1]

The story of entailment, i.e., *relevant* strict implication, in the 20th century is in some respects similar, but in other respects discernibly different.[2] Theories of entailment were, like theories of strict implication, first put forward syntactically, by logicians such as Ackermann (1956) and Anderson and Belnap (1962). Semantic treatments of these and related theories followed in the work of Urquhart (1972), Routley and Meyer (1972, 1973), and Fine (1974). Meyer (1966, 1968) and Routley and Meyer (1972) initiated a program, partially analogous to that of Lemmon (1957), of "refounding" the Anderson-Belnap system **E** of entailment in a theory of relevant modal logic (via translation).[3] However, the program was only a partial success and was not, so far as I am aware, ever extended to other theories of entailment and corresponding relevant modal theories.

This chapter is devoted to a more systematic study of theories of entailment as founded in relevant modal theories than has been carried out to date. I take the work of Meyer (1966, 1968) concerning fragments of **E** and the relevant modal logic **NR** as a jumping off point and consider theories of entailment qua necessitated relevant implication arising from modifying the specifically relevant and/or specifically modal features of relevant modal logics. In this way, I aim to provide newish foundations for theories of entailment; the -ish is intended to indicate that much of what is done here is anticipated by the work of Meyer and also that many of the techniques and constructions employed below are preexisting, if not used to the systematic ends they are put to here.

This is a somewhat ambitious project and I intend to restrict the focus of this chapter along certain dimensions in order to make it manageable. The principal

[1] The history of the development of modal logic and, in particular, its semantics, is not the focus of this chapter and I hope I will be pardoned for being overly brief here. An exceptionally thorough account of the development of the relational (Kripke) semantics for modal logic can be found in Goldblatt (2003, Sect. 4). As Goldblatt (2003, p. 341) emphasizes, while the importance of a binary relation to modeling modality occurred independently to various logicians, Kripke was the first to effectively utilize arbitrary set-theoretic structures as models and so develop a proper model theory for modal logic.

[2] I am here bracketing so-called "analytic" theories of entailment, e.g., those of Parry (1968, pp. 150–153) and Angell (1989). At least in the case of Parry's theory, though, a very similar story can be told. Parry's analytic implication, first developed syntactically, was conjectured by Gödel (1933/1986b) to be analyzable into two parts, basically consisting of a conditional and a filter. Fine (1986) proved this conjecture correct by showing that Parry's implication is basically **S4** strict implication with a filter. In so doing, he also provided a semantics, proved completeness, and separated out the specifically modal and non-modal parts of Parry's theory. For a discussion of various theories of analytic implication and their history, consult Ferguson (2017).

[3] It should be noted that the history of relevant modal logic is older than the history of entailment (at least in the tradition of Ackermann, 1956). Orlov (1928) axiomatized the implication-negation fragment of **R** and also investigated an extension of it with **S4** modal principles. For a discussion, see Došen (1992).

restriction is that I will focus almost entirely on "pure relevant modal logics", that is, relevant modal logics formulated in a language consisting only of the connectives $\{\rightarrow, \Box\}$. While this is undoubtably a restriction, it is in this fragment that almost all the action is, so to speak.

Secondly, I will focus only on a few prominent *normal* relevant modal logics, i.e., relevant modal logics whose specifically modal axioms are at least those of the well-known modal logic **K**.[4] It will, however, be clear from what is done how to extend certain of these results to various other relevant modal logics.

Thirdly, I will consider only two distinct relevant bases for these relevant modal theories: $\mathbf{BCI}_{\rightarrow}$ (alias $\mathbf{RW}_{\rightarrow}$) and $\mathbf{R}_{\rightarrow}$. $\mathbf{BCI}_{\rightarrow}$ is a natural bedrock, in the sense that it is, as seen from a certain perspective, the weakest pure relevant theory of implication proper (as contrasted with entailment).[5]

Fourthly, and finally, I will not primarily be interested in examining connections between previously proposed theories of entailment and relevant modal theories (as was Meyer's interest), but instead in examining the theories of entailment that arise from certain relevant modal theories on their own terms. This notwithstanding, I will make some novel observations concerning the notable theory of entailment $\mathbf{T\text{-}W}_{\rightarrow}$ and a relevant modal theory developed below.

The plan of the chapter is as follows. In Sect. 15.2, I consider desiderata that have been proposed for theories of entailment and present various relevant modal logics axiomatically. A number of syntactic results and independence proofs using matrices are given. I provide a Fine-Urquhart style semantics for these systems in Sect. 15.3 (this sort of semantic framework proves convenient for the systems I am interested in). The semantics is used to establish nontrivial admissibility results for certain weak relevant modal theories. Finally, concluding remarks and suggestions for further research are given in Sect. 15.4.

## 15.2 Relevant Modal Theories and Theories of Entailment

I begin this section by briefly describing the language(s) of the logics I will be discussing in this chapter. I will principally be interested in systems formulated in a language consisting of a countable set of propositional variables $\Pi$, the unary connective $\Box$, and the binary connective $\rightarrow$ (the formation rules are entirely standard). I will also be interested in some systems formulated in the pure implicational language and will annotate them in the usual way (e.g., $\mathbf{L}_{\rightarrow}$). Finally, I will introduce and discuss some systems formulated in a language containing the binary connective

[4] Throughout this chapter, **K** is to be taken in its modal sense and not in the sense of the weakening axiom, viz., $\phi \rightarrow (\psi \rightarrow \phi)$.

[5] I have in mind a proof-theoretic motivation here: $\mathbf{R}_{\rightarrow}$ is what one gets by throwing out weakening from the standard sequent calculus formulation for the pure implicational fragment of intuitionistic logic, and $\mathbf{BCI}_{\rightarrow}$ is what one gets when contraction is thrown out as well. The only structural rule left to eliminate, then, is permutation, but getting rid of permutation yields a theory of entailment (anyway, a theory which satisfies a number of the usual desiderata of a theory of entailment).

$\wedge$ for narrowly technical reasons towards the end of this section. I use $p, q$, etc. for arbitrary propositional variables and $\phi, \psi$, etc. for arbitrary formulae.

Before turning to a discussion of what a theory of entailment is supposed to amount to, it will be convenient to first introduce, for the purpose of reference, a number of axiom schemata and rules:

$$\phi \to \phi \qquad \textbf{(I)}$$

$$(\phi \to \psi) \to ((\chi \to \phi) \to (\chi \to \psi)) \qquad \textbf{(B)}$$

$$(\phi \to \psi) \to ((\psi \to \chi) \to (\phi \to \chi)) \qquad \textbf{(B')}$$

$$(\phi \to (\psi \to \chi)) \to (\psi \to (\phi \to \chi)) \qquad \textbf{(C)}$$

$$(\phi \to (\phi \to \psi)) \to (\phi \to \psi) \qquad \textbf{(W)}$$

$$((\phi \to \phi) \to \psi) \to \psi \qquad \textbf{(SA)}$$

$$\Box(\phi \to \psi) \to (\Box\phi \to \Box\psi) \qquad \textbf{(K)}$$

$$\Box\phi \to \phi \qquad \textbf{(T)}$$

$$\Box\phi \to \Box\Box\phi \qquad \textbf{(4)}$$

$$\frac{\phi, \phi \to \psi}{\psi} \qquad \textbf{(MP)}$$

$$\frac{\phi}{\Box\phi} \qquad \textbf{(NEC)}$$

Now, entailment, in the relevance logic tradition, is understood as necessitated relevant implication. Thus, an entailment connective $\Rightarrow$ can be *defined* in a given system of relevant modal logic in the obvious way: $\phi \Rightarrow \psi := \Box(\phi \to \psi)$. Various desiderata have been proposed for theories of entailment. I list (without necessarily endorsing) some such desiderata below.

**Table 15.1** Systems

| System name | Axioms/Rules |
|---|---|
| **T-W$_{\rightarrow}$** | **B**, **B'**, **I**, **MP** |
| **E$_{\rightarrow}$** | **B'**, **W**, **SA**, **MP** |
| **BCI$^{K4}$** | **B**, **C**, **I**, **K**, **4**, **MP**, **NEC** |
| **R$^{K4}$** | **B**, **C**, **I**, **W**, **K**, **4**, **MP**, **NEC** |
| **BCI$^{S4}$** | **B**, **C**, **I**, **K**, **T**, **4**, **MP**, **NEC** |
| **R$^{S4}$** | **B**, **C**, **I**, **W**, **K**, **T**, **4**, **MP**, **NEC** |

1. Entailment should satisfy transitivity properties (e.g., **B** and **B'**)
2. Any proposition should entail itself (i.e., entailment should satisfy **I**)
3. Entailment should satisfy modus ponens
4. Entailment should satisfy the Ackermann property (i.e., no entailment/necessitative should be entailed by a non-entailment/non-necessitative)
5. Entailment should satisfy the converse Ackermann property (i.e., no entailment/necessitative should entail a non-entailment/non-necessitative)
6. Entailment should satisfy the variable sharing property (i.e., if an entailment is valid, the antecedent and consequent should have a propositional variable in common)

Both the Ackermann property and the variable sharing property have been endorsed by Anderson and Belnap (1975, Sects. 5.1.2, 5.2.1, pp. 33–40) as (at least) being necessary conditions for a theory of entailment (I will specify more formally what these conditions amount to in the present context later). Anderson and Belnap (1975, Sect. 8.12, pp. 95–96) also consider, though do not endorse, a form of the converse Ackermann property.

That an entailment connective should satisfy **B**, **B'**, **I**, and **MP** seems to be widely taken for granted; these properties do, after all, constitute an axiomatic basis for the system **T-W$_{\rightarrow}$**, which has been regarded as a sort of minimal theory of entailment (Anderson & Belnap, 1975, Sect. 8.11, p. 94). On this point, however, at least one prominent exception must be noted: the "syllogistic" relevance logic **S** of Martin and Meyer (1982), which dispenses with **I**.

Table 15.1 lists the main systems I will be discussing. The axiomatization for **E$_{\rightarrow}$** is **E$_{\rightarrow 2}$** from Anderson and Belnap (1975, Sect. 8.3.3, p. 77). The axiomatization for **T-W$_{\rightarrow}$** is that given for it by Anderson and Belnap (1975, Sect. 8.11, p. 94). **R$^{S4}$** is just the (putative) implication-box fragment of the system **NR** from Meyer (1968, p. 475). **R$^{S4}$** and other relevant modal systems from Table 15.1 are also addressed at least en passant (though not in this fragment of the language) by Fuhrmann (1990) and Mares (1993).

I will use the notation **L$^{\Box}$** to refer indifferently to any of the modal relevant systems (i.e., systems listed in Table 15.1 containing modal axioms). Figure 15.1 depicts the inclusion relations among the relevant modal systems from Table 15.1. That these inclusions hold is immediate from their axiomatizations; that they are

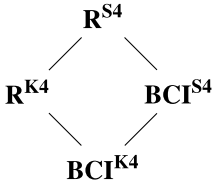



**Fig. 15.1** Inclusion relations among relevant modal systems

proper inclusions is not difficult to show using semantic methods (see Sect. 15.3) or matrices.

Much of this chapter will be devoted to proving certain *admissibility* results for these relevant modal logics. A rule which is admissible in all of the relevant modal logics from Table 15.1 and which I will make extensive use of in the sequel is **RM**:[6]

$$\frac{\phi \to \psi}{\Box\phi \to \Box\psi} \qquad (\mathbf{RM})$$

**Proposition 15.2.1** ***RM*** *is admissible in* $\boldsymbol{BCI^{K4}}$ *(and its extensions).*

***Proof*** Suppose $\vdash \phi \to \psi$. Then we have a derivation of $\Box\phi \to \Box\psi$:

| | | |
|---|---|---|
| 1 | $\phi \to \psi$ | $\vdash$ |
| 2 | $\Box(\phi \to \psi)$ | **NEC** 1 |
| 3 | $\Box(\phi \to \psi) \to (\Box\phi \to \Box\psi)$ | **K** |
| 4 | $\Box\phi \to \Box\psi$ | **MP** 2, 3 |

□

I turn now to considering which properties $\Rightarrow$ satisfies in each of the relevant modal theories catalogued in Table 15.1. First, I will show that each of the relevant modal systems satisfies a form of the Ackermann property. In the context of $\mathbf{E}_{\to}$, with its limited expressive resources, the Ackermann property amounts to the claim that $\nvdash_{\mathbf{E}_{\to}} \phi \to (\psi \to \theta)$ whenever $\phi$ is a propositional variable, for propositional variables are the only non-entailments in the language (Anderson & Belnap, 1975, Sect. 5.2.1, pp. 37–40). In the context of relevant modal theories, which are formulated in a richer language, a somewhat more robust definition of the property is desired.

**Definition 15.2.2** A system of relevant modal logic $\mathbf{L}^{\Box}$ satisfies the *Ackermann property* if $\nvdash_{\mathbf{L}^{\Box}} \phi \to \Box\psi$ whenever $\phi$ is $\Box$-free.

That $\mathbf{R}^{\mathbf{S4}}$ and its subsystems have the Ackermann property, so-defined, is already entailed by a theorem of Meyer (1968, p. 476). However, a somewhat simpler proof can be given using smaller matrices from Mares (1992, pp. 12–13).

**Proposition 15.2.3** $\boldsymbol{R^{S4}}$ *(and each of its subtheories) satisfies the Ackermann property.*

[6] **RM** is not to be confused with the quasi-relevant logic **R**-Mingle. **RM** is named as such following Chellas (1980, p. 17).

***Proof*** Consider the following tables for $\rightarrow$ and $\Box$ where 1 and 2 are designated.

| $\rightarrow$ | 2 | 1 | 0 |
|---|---|---|---|
| 2 | 2 | 0 | 0 |
| 1 | 2 | 1 | 0 |
| 0 | 2 | 2 | 2 |

| $\Box$ | |
|---|---|
| 2 | 1 |
| 1 | 1 |
| 0 | 0 |

A valuation $\nu : \Pi \rightarrow \{0, 1, 2\}$ is extended to the whole language using these matrices. It is tedious but uncomplicated to verify that if $\vdash_{\mathbf{R^{S4}}} \phi$, then $\phi$ takes a designated value on any valuation $\nu$.

Now suppose for contradiction that $\vdash_{\mathbf{R^{S4}}} \phi \rightarrow \Box\psi$ where $\phi$ is $\Box$-free. Consider the valuation $\nu$ such that, for every $p$, $\nu(p) = 2$. By inspection, $\nu(\phi) = 2$ and $\nu(\Box\psi) \neq 2$, hence $\phi \rightarrow \Box\psi$ takes the undesignated value 0 on $\nu$, which is impossible. $\Box$

It is of course an immediate consequence of Proposition 15.2.3 that $\nvdash_{\mathbf{L}^{\Box}} \phi \rightarrow (\psi \Rightarrow \theta)$ whenever $\phi$ is $\Box$-free, and so, a fortiori, whenever $\phi$ is a propositional variable. Indeed, by a straightforward extension of the proof of Proposition 15.2.3, it clearly also follows that $\nvdash_{\mathbf{L}^{\Box}} \phi \Rightarrow (\psi \Rightarrow \theta)$ whenever $\phi$ is $\Box$-free. So in this sense, $\Rightarrow$ also satisfies the Ackermann property as stated for $\mathbf{E}_{\rightarrow}$.

**Definition 15.2.4** A system of relevant modal logic $\mathbf{L}^{\Box}$ satisfies the *converse Ackermann property* if $\nvdash_{\mathbf{L}^{\Box}} \Box\phi \rightarrow \psi$ whenever $\psi$ is $\Box$-free.

It is obvious that neither $\mathbf{BCI^{S4}}$ nor $\mathbf{R^{S4}}$ satisfy the converse Ackermann property, since both have $\mathbf{T}$ as an axiom. However, the removal of $\mathbf{T}$ suffices to yield relevant modal theories which do satisfy the converse Ackermann property, as Proposition 15.2.5 shows.

**Proposition 15.2.5** ***R^K4*** *(and each of its subtheories) satisfies the converse Ackermann property.*

***Proof*** Consider the following tables for $\rightarrow$ and $\Box$ where 1 and 2 are designated.

| $\rightarrow$ | 2 | 1 | 0 |
|---|---|---|---|
| 2 | 2 | 0 | 0 |
| 1 | 2 | 1 | 0 |
| 0 | 2 | 2 | 2 |

| $\Box$ | |
|---|---|
| 2 | 2 |
| 1 | 2 |
| 0 | 2 |

As before, it can be verified that if $\vdash_{\mathbf{R^{K4}}} \phi$, then $\phi$ takes a designated value on any valuation $\nu$. Consider the valuation $\nu$ such that, for every $p$, $\nu(p) = 1$. Then clearly $\nu(\Box\phi) = 2$ and $\nu(\psi) = 1$ where $\psi$ is $\Box$-free, whence $\nu$ falsifies $\Box\phi \rightarrow \psi$. $\Box$

A system which satisfies the Ackermann property "left-directionally" isolates necessitatives and entailments from non-necessitatives and non-entailments, whereas a system which satisfies the converse Ackermann property "right-directionally" isolates them. A system which satisfies both properties (e.g., $\mathbf{R^{K4}}$) effectively sequesters

necessitatives and entailments from non-necessitatives and non-entailments, to the extent licensed by Propositions 15.2.3 and 15.2.5.[7]

I turn now to showing that all of the relevant modal theories discussed here satisfy the variable sharing property. Let $\Pi(\phi)$ be the set of propositional variables occurring in the formula $\phi$.

**Definition 15.2.6** A system of relevant modal logic $\mathbf{L}^{\Box}$ satisfies the *variable sharing property* if $\nvdash_{\mathbf{L}^{\Box}} \phi \to \psi$ whenever $\Pi(\phi) \cap \Pi(\psi) = \emptyset$.

**Proposition 15.2.7** ***R*$^{S4}$** *(and each of its subtheories) satisfies the variable sharing property.*

***Proof*** Recalling that $\mathbf{R}_{\to}$ has the variable sharing property,[8] the easiest way to prove this is to observe that if $\vdash_{\mathbf{R}^{S4}} \phi$, then $\vdash_{\mathbf{R}_{\to}} \epsilon(\phi)$, where $\epsilon$ is a $\Box$-erasing translation: $\epsilon(p) = p$, $\epsilon(\phi \to \psi) = \epsilon(\phi) \to \epsilon(\psi)$, and $\epsilon(\Box\phi) = \epsilon(\phi)$.[9] Supposing that $\Pi(\phi) \cap \Pi(\psi) = \emptyset$ and $\vdash_{\mathbf{L}^{\Box}} \phi \to \psi$, this yields the impossibility that $\vdash_{\mathbf{R}_{\to}} \epsilon(\phi) \to \epsilon(\psi)$. □

So much for the negative properties of these relevant modal logics, and onto their positive properties.

**Proposition 15.2.8** ***BCI*$^{K4}$** *(and each of its extensions) satisfy* ***B***, ***B'***, *and* ***I*** *for entailment.*

***Proof*** It must be proved that $\mathbf{BCI}^{\mathbf{K4}}$ satisfies each of these schemata for $\Rightarrow$. For **I**, $\phi \Rightarrow \phi$ follows immediately from **I** for $\to$ and **NEC**. The proofs of **B** and **B'** are largely the same, so I give just the argument for **B'** here. It must be proved that $(\phi \Rightarrow \psi) \Rightarrow ((\psi \Rightarrow \theta) \Rightarrow (\phi \Rightarrow \theta))$, i.e., $\Box(\Box(\phi \to \psi) \to \Box(\Box(\psi \to \theta) \to \Box(\phi \to \theta)))$.

| | | |
|---|---|---|
| 1 | $(\phi \to \psi) \to ((\psi \to \theta) \to (\phi \to \theta))$ | **B'** |
| 2 | $\Box(\phi \to \psi) \to \Box((\psi \to \theta) \to (\phi \to \theta))$ | **RM** 1 |
| 3 | $\Box((\psi \to \theta) \to (\phi \to \theta)) \to (\Box(\psi \to \theta) \to \Box(\phi \to \theta))$ | **K** |
| 4 | $(2) \to ((3) \to (\Box(\phi \to \psi) \to (\Box(\psi \to \theta) \to \Box(\phi \to \theta))))$ | **B'** |
| 5 | $(3) \to (\Box(\phi \to \psi) \to (\Box(\psi \to \theta) \to \Box(\phi \to \theta)))$ | **MP** 2, 4 |
| 6 | $\Box(\phi \to \psi) \to (\Box(\psi \to \theta) \to \Box(\phi \to \theta))$ | **MP** 3, 5 |
| 7 | $\Box\Box(\phi \to \psi) \to \Box(\Box(\psi \to \theta) \to \Box(\phi \to \theta))$ | **RM** 6 |
| 8 | $\Box(\phi \to \psi) \to \Box\Box(\phi \to \psi)$ | **4** |
| 9 | $(8) \to ((7) \to (\Box(\phi \to \psi) \to \Box(\Box(\psi \to \theta) \to \Box(\phi \to \theta))))$ | **B'** |
| 10 | $(7) \to (\Box(\phi \to \psi) \to \Box(\Box(\psi \to \theta) \to \Box(\phi \to \theta)))$ | **MP** 8, 9 |
| 11 | $\Box(\phi \to \psi) \to \Box(\Box(\psi \to \theta) \to \Box(\phi \to \theta))$ | **MP** 7, 10 |
| 12 | $\Box(\Box(\phi \to \psi) \to \Box(\Box(\psi \to \theta) \to \Box(\phi \to \theta)))$ | **NEC** 11 |

□

[7] Incidentally, there are certain interpretations of $\Box$ where such a thorough separation might be viewed as desirable (for example, consider a deontic interpretation of $\Box$), though I do not myself see why it would be a desirable feature for entailment, properly understood.

[8] For a matrix proof of this, see Anderson and Belnap (1975, Sect. 5.1.2, pp. 33–34). For a semantic proof, see Weiss (2019).

[9] Erasures have a number of uses in modal logic; see, e.g., Chellas (1980, pp. 22–23).

Proposition 15.2.8 gives something of a post facto justification for my focus on relevant modal logics with modal postulates at least as strong as **K4**: all of those modal principles play a role in deriving the transitivity postulates for $\Rightarrow$ in theorem-form. I note that, in the absence of some of these modal principles, the transitivity postulates for $\Rightarrow$ in *rule*-form can still be derived.[10]

**Proposition 15.2.9** ***BCI***$^{S4}$ *(and each of its extensions) satisfy* ***SA*** *and* ***MP*** *for entailment.*

***Proof*** To prove **SA** for $\Rightarrow$, it must be proved that $((\phi \Rightarrow \phi) \Rightarrow \psi) \Rightarrow \psi$, i.e., $\Box(\Box(\Box(\phi \to \phi) \to \psi) \to \psi)$.

| | | |
|---|---|---|
| 1 | $\phi \to \phi$ | **I** |
| 2 | $\Box(\phi \to \phi)$ | **NEC** 1 |
| 3 | $\Box(\phi \to \phi) \to ((\Box(\phi \to \phi) \to \psi) \to \psi)$ | $\vdash$ |
| 4 | $(\Box(\phi \to \phi) \to \psi) \to \psi$ | **MP** 2, 3 |
| 5 | $\Box(\Box(\phi \to \phi) \to \psi) \to \Box\psi$ | **RM** 4 |
| 6 | $\Box\psi \to \psi$ | **T** |
| 7 | $(6) \to ((5) \to (\Box(\Box(\phi \to \phi) \to \psi) \to \psi))$ | **B** |
| 8 | $(5) \to (\Box(\Box(\phi \to \phi) \to \psi) \to \psi)$ | **MP** 6, 7 |
| 9 | $\Box(\Box(\phi \to \phi) \to \psi) \to \psi$ | **MP** 5, 8 |
| 10 | $\Box(\Box(\Box(\phi \to \phi) \to \psi) \to \psi)$ | **NEC** 9 |

To prove **MP** for $\Rightarrow$, it must be shown that if $\vdash \Box(\phi \to \psi)$ and $\vdash \phi$, then $\vdash \psi$. So suppose $\vdash \Box(\phi \to \psi)$ and $\vdash \phi$. Then:

| | | |
|---|---|---|
| 1 | $\Box(\phi \to \psi)$ | $\vdash$ |
| 2 | $\Box(\phi \to \psi) \to (\phi \to \psi)$ | **T** |
| 3 | $\phi \to \psi$ | **MP** 1, 2 |
| 4 | $\phi$ | $\vdash$ |
| 5 | $\psi$ | **MP** 3, 4 |

□

Let $\Phi_{\to}$ be the set of formulae formulated in the pure implicational language and let $\Phi_{\Box,\to}$ be the set of formulae formulated in the language with $\Box$ and $\to$.

**Definition 15.2.10** The *Meyer modal translation* is the function $\mu : \Phi_{\to} \to \Phi_{\Box,\to}$ defined thus:

1. $\mu(p) = p$
2. $\mu(\phi \to \psi) = \Box(\mu(\phi) \to \mu(\psi))$

**Theorem 15.2.11** *For* $\phi \in \Phi_{\to}$, *if* $\vdash_{\boldsymbol{E}_{\to}} \phi$, *then* $\vdash_{\boldsymbol{R}^{S4}} \mu(\phi)$.

[10] The reader might consult Mares and Standefer (2017) for some theories of entailment in the neighborhood of $\mathbf{E}_{\to}$ which only satisfy the transitivity postulates in rule-form. I will leave to future work the interesting project of seeing whether any of these theories of entailment can be captured in weaker relevant modal logics exactly under translation (see Definition 15.2.10 and the remarks on it below).

***Proof*** As $\mu(\phi \to \psi) = \mu(\phi) \Rightarrow \mu(\psi)$, the result follows easily from Propositions 15.2.8 and 15.2.9, together with a proof that **W** holds for $\Rightarrow$. It must be proved that $(\phi \Rightarrow (\phi \Rightarrow \psi)) \Rightarrow (\phi \Rightarrow \psi)$, i.e., $\Box(\Box(\phi \to \Box(\phi \to \psi)) \to \Box(\phi \to \psi))$.

| | | |
|---|---|---|
| 1 | $\Box(\phi \to \psi) \to (\phi \to \psi)$ | **T** |
| 2 | $(\Box(\phi \to \psi) \to (\phi \to \psi)) \to ((\phi \to \Box(\phi \to \psi)) \to (\phi \to (\phi \to \psi)))$ | **B** |
| 3 | $(\phi \to \Box(\phi \to \psi)) \to (\phi \to (\phi \to \psi))$ | **MP** 1, 2 |
| 4 | $(\phi \to (\phi \to \psi)) \to (\phi \to \psi)$ | **W** |
| 5 | $(3) \to ((4) \to ((\phi \to \Box(\phi \to \psi)) \to (\phi \to \psi)))$ | **B'** |
| 6 | $(4) \to ((\phi \to \Box(\phi \to \psi)) \to (\phi \to \psi))$ | **MP** 3, 5 |
| 7 | $(\phi \to \Box(\phi \to \psi)) \to (\phi \to \psi)$ | **MP** 4, 6 |
| 8 | $\Box(\phi \to \Box(\phi \to \psi)) \to \Box(\phi \to \psi)$ | **RM** 7 |
| 9 | $\Box(\Box(\phi \to \Box(\phi \to \psi)) \to \Box(\phi \to \psi))$ | **NEC** 8 |

□

Theorem 15.2.11 is due to Meyer (1966, 1968). Indeed, Meyer (1966, pp. 162–164) showed both Theorem 15.2.11 and its converse for $\mathbf{E}_{\to}$ and $\mathbf{R}^{\mathbf{S4}}$ extended with relevant negation. It was proved by Maksimova (1973) that the result does not extend to the full systems **E** and **NR** (which contains $\mathbf{R}^{\mathbf{S4}}$ as a fragment).

By virtue of Proposition 15.2.8, the reader may have surmised that a version of Theorem 15.2.11 must be in the offing for $\mathbf{T\text{-}W}_{\to}$ and $\mathbf{BCI}^{\mathbf{K4}}$. Indeed, such a result does hold, but it must first be proved that **MP** for $\Rightarrow$ is admissible in $\mathbf{BCI}^{\mathbf{K4}}$. This is, however, a nontrivial result, which will have to await model theoretic tools developed in Sect. 15.3.

Before closing this section, I briefly consider extensions of the relevant modal theories developed above with extensional conjunction ($\wedge$). The following axioms and rule may be added (collectively) to any of the relevant modal theories $\mathbf{L}^{\Box}$:

$$(\phi \wedge \psi) \to \phi \qquad (\wedge\text{E1})$$

$$(\phi \wedge \psi) \to \psi \qquad (\wedge\text{E2})$$

$$((\phi \to \psi) \wedge (\phi \to \theta)) \to (\phi \to (\psi \wedge \theta)) \qquad (\wedge\text{I})$$

$$\Box\phi \wedge \Box\psi \to \Box(\phi \wedge \psi) \qquad (\Box\wedge)$$

$$\frac{\phi, \psi}{\phi \wedge \psi} \qquad (\text{ADJ})$$

Although conjunctive extensions of relevant modal theories are not the particular concern of this chapter, having conjunction around seems to be required (or, anyway, useful) for proving the completeness of $\mathbf{R}_{\to}$-based modal logics (at least, when using the technique used below). Thus, in what follows, one may move indifferently

between $\mathbf{BCI}_{\rightarrow}$-based relevant modal logics containing or not containing conjunction, omitting the conjunction-specific material as desired; however, at least for the sake of proving completeness for $\mathbf{R}^{\mathbf{K4}}$ and $\mathbf{R}^{\mathbf{S4}}$, these will be taken to be formulated with the conjunction axioms given above. Nevertheless, in what follows, I refer to the systems containing conjunction and not indifferently using $\mathbf{L}^{\Box}$, leaving to context and the aforementioned technical note to disambiguate as needed.

## 15.3 Semantics for Relevant Modal Theories

Urquhart (1972, 1973) showed that fragments of various well-known relevance logics could be characterized using operational semantics. Fine (1974) independently developed a hybrid operational-partial order semantics for relevance logic. In this section, I develop a Fine-Urquhart style semantics for the relevant modal systems discussed in Sect. 15.2.[11]

Since my main interest here is in a certain technical application of the semantics (viz., to proving the admissibility of modus ponens for entailment), I will not dwell long on its intuitive interpretation or motivation. I note that Urquhart (1972), in particular, has articulated an informational interpretation of related semantics which might (at least partially) be extended to the present framework.[12] Fine (1974, pp. 348–349, 359) developed a somewhat formalistic account of his semantics in terms of theories, their (relative) closures, and their commitments, and most or all of what he says by way of motivation could also be used to motivate what follows (cf. Logan, 2022).

The structure of this section is as follows. I begin by reviewing a Fine-Urquhart style semantics which suffices for fragments of the logics **BCI** and **R** and then proceed to extend this to the relevant modal systems canvassed in Sect. 15.2. Proofs of soundness and completeness are given and the semantics is used to prove the admissibility of modus ponens for entailment for $\mathbf{BCI}^{\mathbf{K4}}$ and $\mathbf{R}^{\mathbf{K4}}$.

**Definition 15.3.1** A *partially ordered Abelian monoid* (poam, for short) is a structure (frame) $\mathfrak{F} = \langle S, 1, \cdot, \leq \rangle$ where $S$ is a set, $1 \in S$, and $\cdot : S \times S \rightarrow S$, in which the following conditions are imposed:

1. $\langle S, 1, \cdot \rangle$ is an Abelian (commutative) monoid
2. $\langle S, \leq \rangle$ is a partial order
3. $\leq$ is monotonic: $x \leq y$ implies $x \cdot z \leq y \cdot z$

[11] Besides Fine (1974), I believe nearly all previous work on semantics for relevant modal logic is in the Routley-Meyer tradition. Consult, for example, Fuhrmann (1990), Mares and Meyer (1993), and Mares (1993).

[12] The basic idea is that frames consist of pieces of information and there is an operation for combining these which obeys certain properties; formulae are then evaluated at pieces of information. The semantic framework to be developed below would then also require a partial order on the pieces of information and a further operation for handling modality. Various things might be said to motivate these (and the properties imposed on them), but I will not pursue this here.

**Definition 15.3.2** A *square-decreasing partially ordered Abelian monoid* (spoam, for short) is a poam $\mathfrak{F} = \langle S, 1, \cdot, \leq \rangle$ satisfying all of the conditions of Definition 15.3.1 and moreover the condition that $x \cdot x \leq x$.

Poams and spoams are structures fit for modeling fragments of **BCI** and **R**, respectively. The reader will note that the square-decreasing condition in Definition 15.3.2 is the semantic counterpart of **W**.

**Definition 15.3.3** A *(s)poam model* is a structure $\mathfrak{M} = \langle \mathfrak{F}, V \rangle$ where $\mathfrak{F}$ is a (s)poam and $V : \Pi \to \mathcal{P}(S)$ subject to the condition that $x \in V(p)$ and $x \leq y$ imply $y \in V(p)$.

Given a (s)poam model $\mathfrak{M} = \langle S, 1, \cdot, \leq, V \rangle$, the truth conditions for complex formulae are given as follows:

1. $\models_x^{\mathfrak{M}} p$ if and only if $x \in V(p)$
2. $\models_x^{\mathfrak{M}} \phi \wedge \psi$ if and only if $\models_x^{\mathfrak{M}} \phi$ and $\models_x^{\mathfrak{M}} \psi$
3. $\models_x^{\mathfrak{M}} \phi \to \psi$ if and only if for all $y \in S$, if $\models_y^{\mathfrak{M}} \phi$, then $\models_{x \cdot y}^{\mathfrak{M}} \psi$

I now consider semantics for modal extensions. The way the modal operator $\Box$ is to be handled comes directly from Fine (1974, pp. 359–360).

**Definition 15.3.4** A *modalized (square-decreasing) partially ordered Abelian monoid* (m-(s)poam, for short) is a structure $\mathfrak{F} = \langle S, 1, \cdot, ^{\oplus}, \leq \rangle$ where $\langle S, 1, \cdot, \leq \rangle$ is a (s)poam and $^{\oplus} : S \to S$ is such that $x \leq y$ implies $x^{\oplus} \leq y^{\oplus}$.[13]

Various conditions can be imposed on $^{\oplus}$ in m-(s)poams. The conditions I will be interested in, and the axiom schemata they correspond to, are listed in Table 15.2. It should be pointed out that examples of natural mathematical structures satisfying these properties are not too difficult to come by. The structure $\langle \mathbb{R}^+ \cup \{0\}, 0, +, \lfloor \cdot \rfloor, \leq \rangle$, where each of the components is fairly self-explanatory, is an m-poam satisfying all of the properties listed in Table 15.2.

Continuing with the development of the semantics, the truth condition for $\Box$ is as follows:

4. $\models_x^{\mathfrak{M}} \Box\phi$ if and only if $\models_{x^{\oplus}}^{\mathfrak{M}} \phi$

M-(s)poam models are defined just as in Definition 15.3.3. Observe that the heredity condition imposed in Definition 15.3.3 generalizes.

[13] Fine (1974) appears to omit the requirement that $x \leq y$ imply $x^{\oplus} \leq y^{\oplus}$ in his semantics; however, it is required for proving soundness (Theorem 15.3.7).

**Table 15.2** Frame conditions and axioms

| Frame property | Axiom schema | Axiom name |
|---|---|---|
| $1 \leq 1^{\oplus}$ | $\vdash \phi / \vdash \Box\phi$ | **NEC** |
| $x^{\oplus} \cdot y^{\oplus} \leq (x \cdot y)^{\oplus}$ | $\Box(\phi \to \psi) \to (\Box\phi \to \Box\psi)$ | **K** |
| $x^{\oplus} \leq x$ | $\Box\phi \to \phi$ | **T** |
| $x^{\oplus} \leq x^{\oplus\oplus}$ | $\Box\phi \to \Box\Box\phi$ | **4** |

**Proposition 15.3.5** *For any formula $\phi$ and any m-(s)poam model $\mathfrak{M}$, if $\models_x^{\mathfrak{M}} \phi$ and $x \leq y$, then $\models_y^{\mathfrak{M}} \phi$.*

***Proof*** For $\phi$ a propositional variable $p$, the result is by Definition 15.3.3. For $\Box$, suppose $\models_x^{\mathfrak{M}} \Box\phi$ and $x \leq y$; then $\models_{x^{\oplus}}^{\mathfrak{M}} \phi$ and $x^{\oplus} \leq y^{\oplus}$, whence $\models_{y^{\oplus}}^{\mathfrak{M}} \phi$ by the induction hypothesis, and so $\models_y^{\mathfrak{M}} \Box\phi$. Finally, suppose that $\models_x^{\mathfrak{M}} \phi \to \psi$ and $x \leq y$, and consider any $z$ such that $\models_z^{\mathfrak{M}} \phi$. Then $\models_{x \cdot z}^{\mathfrak{M}} \psi$ and, by monotonicity, $x \cdot z \leq y \cdot z$, which suffices since $\models_{y \cdot z}^{\mathfrak{M}} \psi$ follows by the induction hypothesis. □

An m-(s)poam is *fit* for a $\mathbf{BCI}_{\to}$-based ($\mathbf{R}_{\to}$-based) relevant modal system if it is an m-(s)poam which satisfies the frame properties given in Table 15.2 corresponding to its modal axiom schemata and rules. A model is fit for a system $\mathbf{L}^{\Box}$ just if its frame is.

**Definition 15.3.6** A formula $\phi$ is valid in an m-(s)poam model $\mathfrak{M}$ if $\models_1^{\mathfrak{M}} \phi$. A formula $\phi$ is *valid* in $\mathbf{L}^{\Box}$ (in symbols, $\models_{\mathbf{L}^{\Box}} \phi$) if $\models_1^{\mathfrak{M}} \phi$ for all models $\mathfrak{M}$ fit for $\mathbf{L}^{\Box}$.

I turn now to proving soundness and completeness (in what follows, I will sometimes suppress subscripts—e.g., writing $\vdash \phi$ for $\vdash_{\mathbf{L}^{\Box}} \phi$—to improve readability).

**Theorem 15.3.7** *If $\vdash_{\mathbf{L}^{\Box}} \phi$, then $\models_{\mathbf{L}^{\Box}} \phi$.*

***Proof*** The argument is routine. One place where the square-decreasing postulate and Proposition 15.3.5 play a critical role should be pointed out. In verifying soundness for relevant modal systems containing **W**, suppose, for contradiction, that $\not\models (\phi \to (\phi \to \psi)) \to (\phi \to \psi)$. Then there is an m-spoam $\mathfrak{M} = \langle S, 1, \cdot, ^{\oplus}, \leq, V \rangle$ such that $x, y \in S$ and $\models_x^{\mathfrak{M}} \phi \to (\phi \to \psi)$, $\models_y^{\mathfrak{M}} \phi$, and $\not\models_{x \cdot y}^{\mathfrak{M}} \psi$. Then $\models_{(x \cdot y) \cdot y}^{\mathfrak{M}} \psi$. Since $y \cdot y \leq y$, $(y \cdot y) \cdot x \leq y \cdot x$ and $(x \cdot y) \cdot y \leq x \cdot y$ by commutativity and associativity. The desired contradiction now follows from Proposition 15.3.5. □

To prove completeness, I construct a canonical model for $\mathbf{L}^{\Box}$. A set of formulae $\Gamma$ is a $\mathbf{L}^{\Box}$ *theory* if the following conditions are satisfied:

1. $\phi \in \Gamma$ and $\psi \in \Gamma$ imply $\phi \wedge \psi \in \Gamma$
2. $\phi \in \Gamma$ and $\vdash \phi \to \psi$ imply $\psi \in \Gamma$

I write $\mathrm{Th}(\Gamma)$ for the smallest theory containing the set of formula $\Gamma$, or just $\mathrm{Th}(\phi)$ if $\Gamma = \{\phi\}$. By $\mathbb{TH}$, I denote the set of all theories. Define $\Gamma \cdot \Delta = \{\psi : \exists \phi \in \Delta(\phi \to \psi \in \Gamma)\}$ and $\Gamma^{\oplus} = \{\phi : \Box\phi \in \Gamma\}$ (these are local definitions).

**Definition 15.3.8** The *canonical model* for $\mathbf{L}^{\Box}$ is the structure $\mathfrak{M}^c = \langle \mathbb{TH}, \mathbf{L}^{\Box}, \cdot, ^{\oplus}, \subseteq, V^c \rangle$ where $V^c(p) = \{\Gamma : p \in \Gamma\}$.

**Lemma 15.3.9** *The structure* $\mathfrak{M}^c = \langle \mathbb{TH}, \boldsymbol{L}^{\Box}, \cdot, ^{\oplus}, \subseteq, V^c \rangle$ *is an m-(s)poam model fit for* $\boldsymbol{L}^{\Box}$.

***Proof*** It is obvious that $\mathbf{L}^{\Box} \in \mathbb{TH}$. Suppose $\Gamma, \Delta \in \mathbb{TH}$; it must be shown that $\Gamma \cdot \Delta \in \mathbb{TH}$. Suppose $\phi \in \Gamma \cdot \Delta$ and $\psi \in \Gamma \cdot \Delta$; then $\exists \theta, \chi \in \Delta$ such that $\theta \to \phi$, $\chi \to \psi \in \Gamma$. Using **B'**, $\wedge$E1, $\wedge$E2, and the fact that $\Gamma$ is a theory, $(\theta \wedge \chi) \to \phi$, $(\theta \wedge \chi) \to \psi \in \Gamma$. By $\wedge$ I and ADJ, $(\theta \wedge \chi) \to (\phi \wedge \psi) \in \Gamma$. Since, using ADJ, $\theta \wedge \chi \in \Delta$, it follows that $\phi \wedge \psi \in \Gamma \cdot \Delta$, as desired. Suppose $\phi \in \Gamma \cdot \Delta$ and $\vdash \phi \to \psi$; then $\exists \theta \in \Delta$ such that $\theta \to \phi \in \Gamma$. By **B**, $\vdash (\theta \to \phi) \to (\theta \to \psi)$, whence $\theta \to \psi \in \Gamma$ and $\psi \in \Gamma \cdot \Delta$, as required.

Suppose $\Gamma \in \mathbb{TH}$; it must be shown that $\Gamma^{\oplus} \in \mathbb{TH}$. Suppose $\phi \in \Gamma^{\oplus}$ and $\vdash \phi \to \psi$. By **RM**, $\vdash \Box\phi \to \Box\psi$, and by definition, $\Box\phi \in \Gamma$. So $\Box\psi \in \Gamma$ (because $\Gamma$ is a theory) and $\psi \in \Gamma^{\oplus}$ (by definition). Now suppose $\phi \in \Gamma^{\oplus}$ and $\psi \in \Gamma^{\oplus}$; then $\Box\phi \in \Gamma$ and $\Box\psi \in \Gamma$, whence $\Box\phi \wedge \Box\psi \in \Gamma$. Since $\Gamma$ is a theory and $\vdash \Box\phi \wedge \Box\psi \to \Box(\phi \wedge \psi)$ (by $\Box\wedge$), $\Box(\phi \wedge \psi) \in \Gamma$, and so $\phi \wedge \psi \in \Gamma^{\oplus}$.

It has to be shown that $\langle \mathbb{TH}, \mathbf{L}^{\Box}, \cdot \rangle$ is an Abelian monoid. The argument is similar to arguments found in Fine (1974) and Humberstone (1988). Ad commutativity: suppose $\phi \in \Gamma \cdot \Delta$; then $\exists \psi \in \Delta$ such that $\psi \to \phi \in \Gamma$. By **I** and **C**, $\vdash \psi \to ((\psi \to \phi) \to \phi)$. So $(\psi \to \phi) \to \phi \in \Delta$ and $\phi \in \Delta \cdot \Gamma$. The converse is symmetric. Ad the identity: suppose $\phi \in \mathbf{L}^{\Box} \cdot \Gamma$; then $\exists \psi \in \Gamma$ such that $\psi \to \phi \in \mathbf{L}^{\Box}$. So, because $\Gamma$ is a theory, $\phi \in \Gamma$. Conversely, suppose $\phi \in \Gamma$; then since $\phi \to \phi \in \mathbf{L}^{\Box}$ by **I**, $\phi \in \mathbf{L}^{\Box} \cdot \Gamma$, as desired. Ad associativity: suppose $\phi \in \Gamma \cdot (\Delta \cdot \Lambda)$. Then $\exists \psi \in \Delta \cdot \Lambda$ such that $\psi \to \phi \in \Gamma$ and $\exists \theta \in \Lambda$ such that $\theta \to \psi \in \Delta$. By **B**, $\vdash (\psi \to \phi) \to ((\theta \to \psi) \to (\theta \to \phi))$ and so $(\theta \to \psi) \to (\theta \to \phi) \in \Gamma$, $\theta \to \phi \in \Gamma \cdot \Delta$, and $\phi \in (\Gamma \cdot \Delta) \cdot \Lambda$. It has been shown that $\Gamma \cdot (\Delta \cdot \Lambda) \subseteq (\Gamma \cdot \Delta) \cdot \Lambda$; it remains to be shown that $(\Gamma \cdot \Delta) \cdot \Lambda \subseteq \Gamma \cdot (\Delta \cdot \Lambda)$. By commutativity and the already shown half of associativity, $(\Gamma \cdot \Delta) \cdot \Lambda = \Lambda \cdot (\Delta \cdot \Gamma) \subseteq (\Lambda \cdot \Delta) \cdot \Gamma = \Gamma \cdot (\Delta \cdot \Lambda)$, as desired.[14]

That $\langle \mathbb{TH}, \subseteq \rangle$ is a partial order is obvious. For monotonicity, suppose $\Gamma \subseteq \Delta$ and suppose that $\phi \in \Gamma \cdot \Lambda$. Then $\exists \psi \in \Lambda$ such that $\psi \to \phi \in \Gamma \subseteq \Delta$, which suffices to show $\phi \in \Delta \cdot \Lambda$ and therefore $\Gamma \cdot \Lambda \subseteq \Delta \cdot \Lambda$. Suppose $\Gamma \subseteq \Delta$ and suppose $\phi \in \Gamma^{\oplus}$; then $\Box\phi \in \Gamma \subseteq \Delta$, so $\Box\phi \in \Delta$ and $\phi \in \Delta^{\oplus}$. Thus, $\Gamma^{\oplus} \subseteq \Delta^{\oplus}$, as desired.

If $\mathbf{L}^{\Box}$ contains **W**, it must be verified that $\Gamma \cdot \Gamma \subseteq \Gamma$. Suppose $\phi \in \Gamma \cdot \Gamma$; then $\exists \psi \in \Gamma$ such that $\psi \to \phi \in \Gamma$. By ADJ, $\psi \wedge (\psi \to \phi) \in \Gamma$. Using **W**, it can be shown that $\vdash (\psi \wedge (\psi \to \phi)) \to \phi$, whence $\phi \in \Gamma$.

If $\mathbf{L}^{\Box}$ contains **NEC**, then $\mathbf{L}^{\Box} \subseteq \mathbf{L}^{\Box\oplus}$. Suppose $\phi \in \mathbf{L}^{\Box}$, i.e., $\vdash \phi$. Then by **NEC**, $\Box\phi \in \mathbf{L}^{\Box}$, whence $\phi \in \mathbf{L}^{\Box\oplus}$.

If $\mathbf{L}^{\Box}$ contains **K**, then $\Gamma^{\oplus} \cdot \Delta^{\oplus} \subseteq (\Gamma \cdot \Delta)^{\oplus}$. Suppose $\phi \in \Gamma^{\oplus} \cdot \Delta^{\oplus}$. Then $\exists \psi \in \Delta^{\oplus}$ such that $\psi \to \phi \in \Gamma^{\oplus}$, so $\Box\psi \in \Delta$ and $\Box(\psi \to \phi) \in \Gamma$. By **K**, $\Box\psi \to \Box\phi \in \Gamma$, whence $\Box\phi \in \Gamma \cdot \Delta$. Thus, $\phi \in (\Gamma \cdot \Delta)^{\oplus}$, as desired.

[14] I am grateful to Lloyd Humberstone for pointing out that full associativity already follows from the foregoing properties.

If $\mathbf{L}^{\Box}$ contains **T**, then $\Gamma^{\oplus} \subseteq \Gamma$. Suppose $\phi \in \Gamma^{\oplus}$; then $\Box\phi \in \Gamma$, so by **T** and the fact that $\Gamma$ is a theory, $\phi \in \Gamma$.

If $\mathbf{L}^{\Box}$ contains **4**, $\Gamma^{\oplus} \subseteq \Gamma^{\oplus\oplus}$. Suppose $\phi \in \Gamma^{\oplus}$; then $\Box\phi \in \Gamma$, so $\Box\Box\phi \in \Gamma$ by **4**. Consequently, $\Box\phi \in \Gamma^{\oplus}$ and so $\phi \in \Gamma^{\oplus\oplus}$.

Lastly, it is trivial to see that the heredity condition is satisfied, for if $\Gamma \in V^c(p)$, i.e., $p \in \Gamma$, and $\Gamma \subseteq \Delta$, then clearly $p \in \Delta$. □

**Lemma 15.3.10** *For any formula $\phi$ and any $\Gamma \in \mathbb{TH}$, $\models_{\Gamma}^{\mathfrak{M}^c} \phi$ if and only if $\phi \in \Gamma$.*

***Proof*** Where $\phi$ is a propositional variable $p$, the result follows simply by the specification of $V^c$ in Definition 15.3.8.

$\models_{\Gamma}^{\mathfrak{M}^c} \Box\psi$ if and only if $\models_{\Gamma^{\oplus}}^{\mathfrak{M}^c} \psi$ if and only if (by the induction hypothesis) $\psi \in \Gamma^{\oplus}$ if and only if $\Box\psi \in \Gamma$.

$\models_{\Gamma}^{\mathfrak{M}^c} \psi \wedge \theta$ if and only if $\models_{\Gamma}^{\mathfrak{M}^c} \psi$ and $\models_{\Gamma}^{\mathfrak{M}^c} \theta$ if and only if (by the induction hypothesis) $\psi \in \Gamma$ and $\theta \in \Gamma$ if and only if (by ADJ, $\wedge$E1, and $\wedge$E2) $\psi \wedge \theta \in \Gamma$.

Suppose $\psi \to \theta \in \Gamma$ and $\models_{\Delta}^{\mathfrak{M}^c} \psi$; by the induction hypothesis, $\psi \in \Delta$, so $\theta \in \Gamma \cdot \Delta$ whence $\models_{\Gamma\cdot\Delta}^{\mathfrak{M}^c} \theta$, again by the induction hypothesis, which suffices to show $\models_{\Gamma}^{\mathfrak{M}^c} \psi \to \theta$. Conversely, suppose $\psi \to \theta \notin \Gamma$ and consider $\Gamma \cdot \mathrm{Th}(\psi)$. If $\theta \in \Gamma \cdot \mathrm{Th}(\psi)$, then there would be an $\eta \in \mathrm{Th}(\psi)$ such that $\eta \to \theta \in \Gamma$. Then $\vdash \psi \to \eta$, whence $\vdash (\eta \to \theta) \to (\psi \to \theta)$ and $\psi \to \theta \in \Gamma$, which is impossible. So $\theta \notin \Gamma \cdot \mathrm{Th}(\psi)$, from which the result follows easily by the induction hypothesis. □

**Theorem 15.3.11** *If $\models_{\mathbf{L}^{\Box}} \phi$, then $\vdash_{\mathbf{L}^{\Box}} \phi$.*

***Proof*** If $\models_{\mathbf{L}^{\Box}} \phi$, then $\models_{\mathbf{L}^{\Box}}^{\mathfrak{M}^c} \phi$ since $\mathfrak{M}^c$ is an m-(s)poam model fit for $\mathbf{L}^{\Box}$ by Lemma 15.3.9. By Lemma 15.3.10, $\phi \in \mathbf{L}^{\Box}$, i.e., $\vdash_{\mathbf{L}^{\Box}} \phi$. □

We are interested in showing the admissibility of modus ponens for entailment in relevant modal theories, and specifically $\mathbf{BCI}^{\mathbf{K4}}$ and $\mathbf{R}^{\mathbf{K4}}$. The argument that will establish this is model-theoretic in nature, and uses a construction which is adapted from a structure Fine (1974, Sect. 7) used to prove the admissibility of gamma for various relevant logics.[15]

**Definition 15.3.12** Let $\mathfrak{M} = \langle S, 1, \cdot, ^{\oplus}, \leq, V\rangle$ be an m-(s)poam in which $1 \neq 1^{\oplus}$. Define the structure $\mathfrak{M}' = \langle S', 1', \cdot', ^{\oplus'}, \leq', V'\rangle$ with a point $1' \notin S$ thus:

1. $S' = S \cup \{1'\}$
2. $x \cdot' y = \begin{cases} x \cdot y \text{ if } x, y \in S, \\ x \text{ if } y = 1', \\ y \text{ if } x = 1'. \end{cases}$
3. $x^{\oplus'} = \begin{cases} x^{\oplus} \text{ if } x \in S, \\ 1' \text{ if } x = 1'. \end{cases}$

[15] In standard relevance logics, gamma, the inference from $\neg\phi \vee \psi$ and $\phi$ to $\psi$, is notoriously invalid though admissible. For a discussion of the history of gamma and proofs of its admissibility, consult Urquhart (2016).

4. $x \leq' y$ if and only if $\begin{cases} x \leq y \text{ if } x, y \in S, \\ 1 \leq y \text{ if } y \in S, x = 1', \\ x = 1' \text{ if } y = 1'. \end{cases}$
5. $x \in V'(p)$ if and only if $\begin{cases} x \in V(p) \text{ if } x \in S, \\ 1 \in V(p) \text{ if } x = 1'. \end{cases}$

**Lemma 15.3.13** *If $\mathfrak{M} = \langle S, 1, \cdot, ^{\oplus}, \leq, V\rangle$ is an m-(s)poam model satisfying at least the frame condition for **NEC** given in Table 15.2, then $\mathfrak{M}' = \langle S', 1', \cdot', ^{\oplus'}, \leq', V'\rangle$ is an m-(s)poam satisfying the same frame conditions.*

***Proof*** The argument for this is mostly tedious case checking. It's straightforward to see $\cdot'$ and $^{\oplus'}$ are operations on $S'$. Showing $\cdot'$ satisfies the requisite properties involves enumerating cases depending on which arguments (if any) equal $1'$. For example, for verifying associativity—$x \cdot' (y \cdot' z) = (x \cdot' y) \cdot' z$—there are eight cases to consider. One is the case in which $x = z = 1'$ and $y \in S$; then on the left hand side we have $x \cdot' (y \cdot' z) = 1' \cdot' (y \cdot' 1') = 1' \cdot' y = y$ and on the right hand side we have $(x \cdot' y) \cdot' z = (1' \cdot' y) \cdot' 1' = y \cdot' 1' = y$, as desired.

To show $\leq'$ is a partial order satisfying monotonicity, it is just more case checking. Ad reflexivity: if $x \in S$, $x \leq' x$ follows from $x \leq x$; otherwise, it holds by definition. Ad antisymmetry: suppose $x \leq' y$ and $y \leq' x$. If $x, y \in S$, then $x \leq y$ and $y \leq x$, whence $x = y$. If $x \notin S$ (i.e., $x = 1'$), either $y \in S$ or $y = 1'$. The first case is impossible because $y \leq' 1'$ only if $y = 1'$. On the second case, we have $x = y = 1'$. Ad transitivity: suppose $x \leq' y$ and $y \leq' z$. If $z = 1'$, then $x = y = 1'$ and $x \leq' z$. So consider cases in which $z \in S$. If $y = 1'$, then $x = 1'$ and $1 \leq z$, and so $x \leq' z$. If $y \in S$, then if $x = 1'$, then $1 \leq y \leq z$, so $1 \leq z$ and $z \in S$ and $x = 1'$, from which it follows that $x \leq' z$. Lastly, the result follows by the transitivity of $\leq$ if $x, y, z \in S$. Ad monotonicity: suppose $x \leq' y$. If $y = 1'$, then $x = 1'$, and it follows that $x \cdot' z \leq' y \cdot' z$ for any $z \in S'$. If $z = 1'$, then again, $x \cdot' z \leq' y \cdot' z$ for any $x, y \in S'$. So suppose $y, z \in S$; either $x \in S$ or $x = 1'$. In the first case, $x \cdot' z \leq' y \cdot' z$ because $x \leq y$ and so $x \cdot z \leq y \cdot z$. In the second, $x = 1'$, $y \in S$, and $x \leq' y$ imply $1 \leq y$ which implies $z \leq y \cdot z$, whence $z \leq' y \cdot' z$ and $x \cdot' z = z \leq' y \cdot' z$, as desired.

It has to be shown that $x \leq' y$ implies $x^{\oplus'} \leq' y^{\oplus'}$. If $y = 1'$, then $x = 1'$, and so the result holds as $x^{\oplus'} = y^{\oplus'} = 1'^{\oplus'} = 1'$. Otherwise, $x = 1'$ or $x \neq 1'$. In the first case, $1 \leq y$ implies $1^{\oplus} \leq y^{\oplus}$; since $1 \leq 1^{\oplus}$ ex hypothesi, $1 \leq y^{\oplus}$, whence $x^{\oplus'} = 1'^{\oplus'} \leq' y^{\oplus} = y^{\oplus'}$, as desired. In the second case, it must be that $x \leq y$, from which the result follows immediately.

If $\mathfrak{M}$ is square-decreasing, $\mathfrak{M}'$ is also: it is to be shown that $x \cdot' x \leq' x$. If $x \in S$, the result follows by the fact that $\mathfrak{M}$ is square-decreasing. If $x = 1'$, it obviously holds that $1' = 1' \cdot' 1' \leq' 1'$.

I turn now to verifying the various modal properties. Since $1'^{\oplus'} = 1'$, it is obviously the case that if $\mathfrak{M}$ satisfies $1 \leq 1^{\oplus}$, $\mathfrak{M}'$ satisfies the corresponding condition.

Suppose $\mathfrak{M}$ satisfies $x^{\oplus} \cdot y^{\oplus} \leq (x \cdot y)^{\oplus}$, to show that $\mathfrak{M}'$ satisfies the corresponding condition. If $x, y \in S$, the result is immediate from the fact that $\mathfrak{M}$ satisfies $x^{\oplus} \cdot y^{\oplus} \leq (x \cdot y)^{\oplus}$. If $x \in S$ and $y = 1'$, then $x^{\oplus'} \cdot' 1'^{\oplus'} = x^{\oplus'} \cdot' 1' = x^{\oplus'} = (x \cdot' 1')^{\oplus'}$, which suffices; the case where $y \in S$ and $x = 1'$ is symmetric. Finally, if $x = y = 1'$, then the result is obvious since $1' \leq' 1'$.

Suppose $\mathfrak{M}$ satisfies $x^{\oplus} \leq x^{\oplus\oplus}$, to show that $\mathfrak{M}'$ satisfies the corresponding condition. If $x \in S$, then $x^{\oplus'} = x^{\oplus} \leq x^{\oplus\oplus} = x^{\oplus'\oplus'}$, so $x^{\oplus'} \leq' x^{\oplus'\oplus'}$. If $x \notin S$, the result is obvious.

Ex hypothesi, $\mathfrak{M}$ does not satisfy the condition that $x^{\oplus} \leq x$, else it would already be the case that $1 = 1^{\oplus}$.

For heredity, suppose $x \in V'(p)$ and $x \leq' y$. If $y = 1'$, then $x = 1'$ and clearly $y = x \in V'(p)$. Otherwise, either $x = 1'$ or not. In the first case, since $1 \leq y$ and $1 \in V(p)$ by the specification of $V'$, $y \in V(p)$ and so $y \in V'(p)$. Alternatively, the result must hold since $x \leq y$ and $x \in V'(p)$ imply $x \in V(p)$ and $y \in V(p)$, whence $y \in V'(p)$. □

**Lemma 15.3.14** *Given* $\mathfrak{M} = \langle S, 1, \cdot, ^{\oplus}, \leq, V\rangle$ *and* $\mathfrak{M}' = \langle S', 1', \cdot', ^{\oplus'}, \leq', V'\rangle$, *for any* $x \in S$: $\models_x^{\mathfrak{M}'} \phi$ *if and only if* $\models_x^{\mathfrak{M}} \phi$.

***Proof*** This is true by Definition 15.3.12 for the case where $\phi$ is a variable $p$. If $\phi$ is a conjunction $\psi \wedge \theta$, then: $\models_x^{\mathfrak{M}'} \psi \wedge \theta$ if and only if $\models_x^{\mathfrak{M}'} \psi$ and $\models_x^{\mathfrak{M}'} \theta$ if and only if (by the induction hypothesis) $\models_x^{\mathfrak{M}} \psi$ and $\models_x^{\mathfrak{M}} \theta$ if and only if $\models_x^{\mathfrak{M}} \psi \wedge \theta$. If $\phi$ is a necessitation $\Box\psi$, then: $\models_x^{\mathfrak{M}'} \Box\psi$ if and only if $\models_{x^{\oplus'}}^{\mathfrak{M}'} \psi$ if and only if (by the induction hypothesis and fact that $x^{\oplus'} = x^{\oplus}$) $\models_{x^{\oplus}}^{\mathfrak{M}} \psi$ if and only if $\models_x^{\mathfrak{M}} \Box\psi$.

Finally, suppose $\phi$ is a conditional $\psi \rightarrow \theta$. If $\models_x^{\mathfrak{M}'} \psi \rightarrow \theta$ and $\models_y^{\mathfrak{M}'} \psi$ (for arbitrary $y \in S$) then, by the induction hypothesis, $\models_y^{\mathfrak{M}} \psi$, whence $\models_{x \cdot' y}^{\mathfrak{M}'} \theta$. So, by the induction hypothesis and fact that $x \cdot' y = x \cdot y$, $\models_{x \cdot y}^{\mathfrak{M}} \theta$, which suffices. Conversely, suppose that $\not\models_x^{\mathfrak{M}'} \psi \rightarrow \theta$; then $\exists y \in S'$ such that $\models_y^{\mathfrak{M}'} \psi$ and $\not\models_{x \cdot' y}^{\mathfrak{M}'} \theta$. If $y \in S$, then by the induction hypothesis and fact that $x \cdot' y = x \cdot y$, $\models_y^{\mathfrak{M}} \psi$ and $\not\models_{x \cdot y}^{\mathfrak{M}} \theta$, from which it follows that $\not\models_x^{\mathfrak{M}} \psi \rightarrow \theta$. Alternatively, $y = 1'$; then $\models_y^{\mathfrak{M}'} \psi$ and $y \leq' 1$ imply $\models_1^{\mathfrak{M}'} \psi$, and $\not\models_{x \cdot' y}^{\mathfrak{M}'} \theta$ and $x \cdot' 1' = x$ imply $\not\models_x^{\mathfrak{M}'} \theta$. By the induction hypothesis, $\models_1^{\mathfrak{M}} \psi$ and $\not\models_x^{\mathfrak{M}} \theta$, whence $\not\models_x^{\mathfrak{M}} \psi \rightarrow \theta$. □

We are finally in a position to prove the admissibility of modus ponens for entailment ($\Rightarrow$) in the relevant modal logics $\mathbf{BCI}^{\mathbf{K4}}$ and $\mathbf{R}^{\mathbf{K4}}$. In fact, this is really a corollary of a more general result concerning the admissibility of the converse of **NEC**:

$$\frac{\Box\phi}{\phi} \qquad \textbf{(cNEC)}$$

**Theorem 15.3.15** ***cNEC*** *is admissible in relevant modal theories for which Lemmata 15.3.13 and 15.3.14 apply, including* $\boldsymbol{BCI^{K4}}$ *and* $\boldsymbol{R^{K4}}$.

***Proof*** Suppose $\vdash \Box\phi$ and $\nvdash \phi$. By Theorems 15.3.7 and 15.3.11, $\models \Box\phi$ but $\not\models \phi$. Then there is an m-(s)poam $\mathfrak{M} = \langle S, 1, \cdot, ^{\oplus}, \leq, V\rangle$ in which $\not\models_1^{\mathfrak{M}} \phi$. Consider the model $\mathfrak{M}' = \langle S', 1', \cdot', ^{\oplus'}, \leq', V'\rangle$ obtained from $\mathfrak{M}$ by Definition 15.3.12. By Lemma 15.3.13, $\mathfrak{M}'$ satisfies the same frame conditions. Moreover, by Lemma 15.3.14, $\not\models_1^{\mathfrak{M}'} \phi$. As $\models_{1'}^{\mathfrak{M}'} \Box\phi$ (ex hypothesi), $\models_{1'^{\oplus'}}^{\mathfrak{M}'} \phi$, and so $\models_{1'}^{\mathfrak{M}'} \phi$ because $1'^{\oplus'} = 1'$. But $1' \leq' 1$ and $\models_{1'}^{\mathfrak{M}'} \phi$ imply $\models_1^{\mathfrak{M}'} \phi$ by Proposition 15.3.5, which is a contradiction. □

**Corollary 15.3.16** *Modus ponens for entailment is admissible in relevant modal theories for which Theorem 15.3.15 applies, including* $\boldsymbol{BCI^{K4}}$ *and* $\boldsymbol{R^{K4}}$.

***Proof*** Suppose $\vdash \Box(\phi \to \psi)$ and $\vdash \phi$ but $\nvdash \psi$. By **cNEC**, licensed by Theorem 15.3.15, $\vdash \phi \to \psi$, whence $\vdash \psi$ since $\vdash \phi$ and all of the systems have **MP**. Contradiction. □

**Theorem 15.3.17** *For* $\phi \in \Phi_{\to}$, *if* $\vdash_{\textbf{T-W}_{\to}} \phi$, *then* $\vdash_{\boldsymbol{BCI^{K4}}} \mu(\phi)$.

***Proof*** Immediate from Proposition 15.2.8 and Corollary 15.3.16. □

It would be of considerable interest to show that the *converse* of Theorem 15.3.17 also obtains: for $\phi \in \Phi_{\to}$, $\vdash_{\mathbf{BCI^{K4}}} \mu(\phi)$ implies $\vdash_{\textbf{T-W}_{\to}} \phi$. Reflection on Meyer's proof of the converse of Theorem 15.2.11 suggests that the proof of such a result would require proof-theoretic techniques beyond the scope of the present chapter.

In connection with $\textbf{T-W}_{\to}$, one other thing should be mentioned. It is known that $\vdash_{\textbf{T-W}_{\to}} \phi \to \psi$ and $\vdash_{\textbf{T-W}_{\to}} \psi \to \phi$ only if $\phi$ and $\psi$ are the same formula.[16] Observe that this is *not* the case for $\Rightarrow$ in $\mathbf{BCI^{K4}}$: $\vdash_{\mathbf{BCI^{K4}}} \Box(\phi \to ((\phi \to \phi) \to \phi))$ and $\vdash_{\mathbf{BCI^{K4}}} \Box(((\phi \to \phi) \to \phi) \to \phi)$, but obviously $\phi$ and $(\phi \to \phi) \to \phi$ are not the same formula.[17]

## 15.4 Concluding Remarks

In this chapter, I presented a systematic study of theories of entailment as developed in several relevant modal logics. I considered desiderata that a theory of entailment ought to satisfy in Sect. 15.2. Taking stock, the relatively weak relevant modal logics $\mathbf{BCI^{K4}}$ and $\mathbf{R^{K4}}$ satisfy all of the criteria considered there, while the stronger relevant modal logics $\mathbf{BCI^{S4}}$ and $\mathbf{R^{S4}}$ satisfy all of the criteria except the converse Ackermann property.

Demonstrating (or refuting) that these systems met the stated properties was mostly straightforward, with the exception of proving the admissibility of modus ponens for entailment for $\mathbf{BCI^{K4}}$ and $\mathbf{R^{K4}}$. This required a somewhat sophisticated semantic argument and, to that end, I presented semantics for each of the relevant modal logics in Sect. 15.3 and proved soundness and completeness.

---

[16] That this should be so is the famous '**P-W** Problem' of Anderson and Belnap (1975, Sect. 8.11, p. 95). The conjecture was proved by Martin and Meyer (1982). For subsequent discussion and alternative proofs of the conjecture, the reader may consult Kron (1985), Komori (1994), and Hirokawa et al. (2000).

[17] Define (cf. Urquhart, 1972, p. 161):

$$\Diamond^k \phi = \begin{cases} \phi & k = 0 \\ (\Diamond^{k-1}\phi \to \Diamond^{k-1}\phi) \to \Diamond^{k-1}\phi & k > 0 \end{cases}$$

Then it can be proved by induction that, for all integers $k \geq 0$, $\vdash_{\mathbf{BCI}_{\to}} \Diamond^k \phi \to \phi$ and $\vdash_{\mathbf{BCI}_{\to}} \phi \to \Diamond^k \phi$.

Some connections between these relevant modal logics and preexisting theories of entailment were noted. In particular, I pointed out that $\mathbf{E}_{\rightarrow}$ can be interpreted in $\mathbf{R}^{\mathbf{S4}}$ under the Meyer translation $\mu$, and the considerably less trivial fact that $\mathbf{T\text{-}W}_{\rightarrow}$ can be interpreted in $\mathbf{BCI}^{\mathbf{K4}}$ under the same translation. Whether or not the translation is exact in the case of $\mathbf{T\text{-}W}_{\rightarrow}$ and $\mathbf{BCI}^{\mathbf{K4}}$ is an open problem.

As I indicated in Sect. 15.1, the scope of this chapter was restricted in various ways. Besides the interesting open problem noted in the previous paragraph, there are many other things concerning the topics of this chapter worth looking into. These include, but are not limited to, investigating the properties of theories of entailment in relevant modal logics with richer languages (e.g., languages including negation and disjunction) and examining the properties of theories of entailment over relevant modal logics with weaker or stronger relevant and/or modal components. I hope this chapter will serve as inspiration for some further work along these lines.

**Acknowledgements** I would like to thank an anonymous referee, Lloyd Humberstone, Edwin Mares, and Graham Priest for comments and/or suggestions. I would also like to thank the series editor, Sven-Ove Hansson, for handling the refereeing of this chapter. Finally, I would like to express my gratitude to Saul Kripke for all that he has done to advance the study of modal and non-classical logic and make research such as this possible.

## References

Ackermann, W. (1956). Begründung einer strengen Implikation. *Journal of Symbolic Logic, 21*(2), 113–128.

Anderson, A. R., & Belnap, N. D. (1962). The pure calculus of entailment. *Journal of Symbolic Logic, 27*(2), 19–52.

Anderson, A. R., & Belnap, N. D. (1975). *Entailment: The logic of relevance and necessity* (Vol. I). Princeton: Princeton University Press.

Angell, R. B. (1989). Deducibility, entailment and analytic containment. In J. Norman & R. Sylvan (Eds.), *Directions in relevant logic* (pp. 119–143). Dordrecht: Kluwer Academic Publishers.

Chellas, B. F. (1980). *Modal logic: An introduction*. New York: Cambridge University Press.

Došen, K. (1992). The first axiomatization of relevant logic. *Journal of Philosophical Logic, 21*(4), 339–356.

Ferguson, T. M. (2017). *Meaning and proscription in formal logic: Variations on the propositional logic of William T. Parry*. Cham: Springer.

Fine, K. (1974). Models for entailment. *Journal of Philosophical Logic, 3*(4), 347–372.

Fine, K. (1986). Analytic implication. *Notre Dame Journal of Formal Logic, 27*(2), 169–179.

Fuhrmann, A. (1990). Models for relevant modal logics. *Studia Logica, 49*(4), 501–514.

Gödel, K. (1933/1986a). Eine interpretation des intuitionistischen aussagenkalküls. In S. Feferman, J. W. Dawson, Jr., S. C. Kleene, G. H. Moore, R. M. Solovay, & J. van Heijenoort (Eds.), *Kurt Gödel: Collected works* (Vol. 1, pp. 300–303). New York: Oxford University Press.

Gödel, K. (1933/1986b). Über die Parryschen axiome. In S. Feferman, J. W. Dawson, Jr., S. C. Kleene, G. H. Moore, R. M. Solovay, & J. van Heijenoort (Eds.), *Kurt Gödel: Collected works* (Vol. 1, pp. 266–267). New York: Oxford University Press.

Goldblatt, R. (2003). Mathematical modal logic: A view of its evolution. *Journal of Applied Logic, 1*(5–6), 309–392.

Hirokawa, S., Komori, Y., & Nagayama, M. (2000). A lambda proof of the P-W theorem. *Journal of Symbolic Logic, 65*(4), 1841–1849.
Humberstone, I. L. (1988). Operational semantics for positive R. *Notre Dame Journal of Formal Logic, 29*(1), 61–80.
Komori, Y. (1994). Syntactical investigations into *BI* logic and *BB'I* logic. *Studia Logica, 53*(3), 397–416.
Kripke, S. A. (1959). A completeness theorem in modal logic. *Journal of Symbolic Logic, 24*(1), 1–14.
Kripke, S. A. (1963). Semantical analysis of modal logic I: Normal modal propositional calculi. *Zeitschrift für mathematische Logik und Grundlagen der Mathematik, 9*(5–6), 67–96.
Kripke, S. A. (1965). Semantical analysis of modal logic II: Non-normal modal propositional calculi. In J. W. Addison, L. Henkin, & A. Tarski (Eds.), *The theory of models, Proceedings of the 1963 international symposium at Berkeley* (pp. 206–220). Amsterdam: North-Holland Publishing Company.
Kron, A. (1985). A constructive proof of a theorem in relevance logic. *Zeitschrift für mathematische Logik und Grundlagen der Mathematik, 31*(25–28), 423–430.
Lemmon, E. J. (1957). New foundations for Lewis modal systems. *Journal of Symbolic Logic, 22*(2), 176–186.
Lewis, C. I., & Langford, C. H. (1932/1959). *Symbolic logic* (2nd ed.). New York: Dover Publications.
Logan, S. A. (2022). Deep fried logic. *Erkenntnis, 87*(1), 257–286.
Maksimova, L. (1973). A semantics for the calculus *E* of entailment. *Bulletin of the Section of Logic, 2*(1), 18–21.
Mares, E. (1992). Andersonian deontic logic. *Theoria, 58*(1), 3–20.
Mares, E. (1993). Classically complete modal relevant logics. *Mathematical Logic Quarterly, 39*(1), 165–177.
Mares, E. & Meyer, R. K. (1993). The semantics of R4. *Journal of Philosophical Logic, 22*(1), 95–110.
Mares, E., & Standefer, S. (2017). The relevant logic E and some close neighbours: A reinterpretation. *IFCoLog Journal of Logics and Their Applications, 4*(3), 695–730.
Martin, E. P., & Meyer, R. K. (1982). Solution to the P-W problem. *Journal of Symbolic Logic, 47*(4), 869–887.
Meyer, R. K. (1966). *Topics in modal and many-valued logic* (Unpublished doctoral dissertation). University of Pittsburgh.
Meyer, R. K. (1968). Entailment and relevant implication. *Logique et Analyse, 11*(44), 472–479.
Orlov, I. E. (1928). The calculus of compatibility of propositions (in Russian). *Matematicheskiĭ Sbornik*, *35*, 263–286.
Parry, W. T. (1968). The logic of C. I. Lewis. In P. A. Schilpp (Ed.), *The philosophy of C. I. Lewis* (pp. 115–154). La Salle, IL: Open Court.
Routley, R., & Meyer, R. K. (1972). The semantics of entailment: II. *Journal of Philosophical Logic, 1*(1), 53–73.
Routley, R., & Meyer, R. K. (1973). The semantics of entailment. In H. Leblanc (Ed.), *Truth, syntax, and modality: Proceedings of the Temple University conference on alternative semantics* (Vol. 68, pp. 199–243). Amsterdam: North-Holland Publishing Company.
Urquhart, A. (1972). Semantics for relevant logics. *Journal of Symbolic Logic, 37*(1), 159–169.
Urquhart, A. (1973). *The semantics of entailment* (Unpublished doctoral dissertation). University of Pittsburgh.
Urquhart, A. (2016). The Story of $\gamma$. In K. Bimbó (Ed.), *J. Michael Dunn on information based logics* (Vol. 8, pp. 93–105). Switzerland: Springer.
Weiss, Y. (2019). A note on the relevance of semilattice relevance logic. *Australasian Journal of Logic, 16*(6), 177–185.

# Chapter 16
# Accepting a Logic, Accepting a Theory

**Timothy Williamson**

**Abstract** This chapter responds to Saul Kripke's critique of the idea of adopting an alternative logic. It defends an anti-exceptionalist view of logic, on which coming to accept a new logic is a special case of coming to accept a new scientific theory. The approach is illustrated in detail by debates on quantified modal logic. A distinction between folk logic and scientific logic is modelled on the distinction between folk physics and scientific physics. The importance of not confusing logic with metalogic in applying this distinction is emphasized. Defeasible inferential dispositions are shown to play a major role in theory acceptance in logic and mathematics as well as in natural and social science. Like beliefs, such dispositions are malleable in response to evidence, though not simply at will. The chapter considers the Quinean objection that accepting an alternative logic involves changing the subject rather than denying the doctrine. The objection is shown to depend on neglect of the social dimension of meaning determination, akin to descriptivism about proper names and natural kind terms criticized by Kripke and Putnam. Normal standards of interpretation indicate that disputes between classical and non-classical logicians are genuine disagreements.

**Keywords** Alternative logics · Dummett · Intuitionistic logic · Kripke · Modal logic · Putnam · Quine

## 16.1 Introduction

I first encountered Saul Kripke in my first term as an undergraduate, studying mathematics and philosophy at Oxford, when he gave the 1973 John Locke Lectures (later published as Kripke, 2013). A huge audience listened intently, including most of the brightest Oxford philosophers of the time. Kripke was a charismatic figure; his head looked as though it belonged on an Assyrian monumental bull, and he gave a

T. Williamson (✉)
Oxford University, Oxford, UK
e-mail: timothy.williamson@philosophy.ox.ac.uk

© The Author(s), under exclusive license to Springer Nature Switzerland AG 2024

Y. Weiss and R. Birman (eds.), *Saul Kripke on Modal Logic*, Outstanding Contributions to Logic 30, https://doi.org/10.1007/978-3-031-57635-5_16

sense of easy intellectual power in reserve. I was struck by his combination of crystal clarity, informal rigour, apt examples, good sense, and pointed humour. He became the closest I had to a model of how to do philosophy. I used to pore over a little book of readings on reference and modality (Linsky, 1971); the chapter to which I most often returned was a reprint of 'Semantical Considerations on Modal Logic' (Kripke, 1963). I raced through 'Naming and Necessity' at a sitting (it was already available as Kripke, 1972). Naturally, in my own book on modal logic, Kripke's work occupies a central role; he figures on sixty of its pages (Williamson, 2013).

This chapter comes at logic, including modal logic, from a different angle, prompted by a different aspect of Kripke's thinking. In 'The Question of Logic' (Kripke, 2020, derived from a lecture originally given in 1974, though I heard about it only decades later), he argues that the common philosophical talk of 'adopting an alternative logic' to solve a problem is fundamentally confused, because without already having logic one cannot derive the consequences of a specification of the 'alternative system'.

By contrast, in the Oxford of my philosophical upbringing, the dominant philosophical logician was Michael Dummett, who supervised me for the final year of my doctoral studies (1979–80). Dummett presented intuitionistic logic, with its rejection of the law of excluded middle, as an alternative logic which one arguably *should* adopt, on grounds in the philosophy of language, and thereby reject classical logic (Dummett, 1975). Under his influence, some of my Oxford contemporaries *had* to all appearances taken the plunge and adopted intuitionistic logic. They certainly thought they had. Had they done the very thing which, for Kripke, makes no sense? I was never at all tempted to join them; with my firmly realist sympathies, I found the philosophical case for intuitionistic logic utterly unconvincing. Nevertheless, I did not doubt my contemporaries when they claimed to have adopted intuitionistic logic.

'The Question of Logic' focusses primarily on a different deviation from classical logic: Hilary Putnam's case for adopting 'quantum logic' to resolve problems in explaining experimental results in quantum mechanics (Putnam, 1968; he retracted the view for other reasons in Putnam, 2005). However, Kripke briefly discusses intuitionistic logic in the paper, rejecting claims that intuitionists have adopted it as an alternative to classical logic.

I strongly agree with Kripke that the cases offered for adopting one or other formal system as an alternative logic are often very superficial, in ways closely related to his critique of adoptionism. One striking example—though not the focus of Kripke's arguments in 'The Question of Logic'—is that advocates of non-classical logics often rely in their metalogical reasoning on classically valid principles underivable in their preferred non-classical system, while providing no serious justification for doing so. This practice can easily make a significant difference. For example, many theorists have turned to various forms of three-valued or continuum-valued semantics to handle sorites paradoxes for vague languages. Such accounts are typically taken to imply that the law of excluded middle is invalid in the object-language. However, by using excluded middle in the metalanguage, one can prove valid some formulas of the object-language which cannot be proved valid without it. Moreover, the phenomenon of higher-order vagueness strongly suggests that many-valued logicians'

vagueness-related reasons for resisting excluded middle in the object-language generalize naturally to excluded middle in the metalanguage. In these circumstances, many-valued logicians' tendency to use classical logic, with excluded middle, in the metalanguage looks problematic. Indeed, if by their own lights they cannot rely on classical logic in the metalanguage, it is quite unclear what logical principles they are entitled to rely on in the metalanguage for purposes of calculating which logical principles are valid in the object-language. They should rely only on those valid for a vague language, but which those are is just what they are trying to work out (Williamson, 1994, pp. 111–13, 127–30). That is reminiscent of the problems Kripke identifies for 'adopting a logic'.

Many critics of classical logic deny that their criticisms extend to its applications to mathematics: allegedly, since the language of pure mathematics is precise and unparadoxical, no revision of standard mathematics is required. But that lazy view ignores applications of pure mathematics to the concrete world, as in physics, which require instantiating variables in the mathematical theorems by terms from outside the language of pure mathematics. Their original criticisms, if sound, apply to the extended language. Thus they are in no position to shirk the hard task of reconstructing as much mathematics as they can on the basis of their proposed non-classical logic (Williamson, 2018). Indeed, in 'The Question of Logic', Kripke gives a delightful argument that Putnam's revisionary claims commit him to holding that $2 \times 2$ is at least 5.

A similar dialectic plays out if the critics of classical logic deny that their criticisms extend to its applications to their model-theoretic semantics, usually formulated in the language of mathematics, specifically set theory, extended by some vocabulary for syntax: allegedly, since such a language is precise and unparadoxical, no revision of their model theory is required. But that lazy view ignores their eventual applications of the model theory to *intended* models for the language, which require instantiating variables in the model theory by terms in an extended language capable of specifying the possibly vague or paradoxical intended readings. Their original criticisms of classical logic, if sound, apply to such a language. Thus they are in no position to shirk the hard task of reconstructing as much of their model theory as they can from the starting-point of their proposed non-classical logic.

However, analogous objections would be unfair to *some* proponents of alternative logics, who *do* restrict themselves to their officially endorsed principles of non-classical logic in the metalanguage too. In particular, there is a well-developed tradition of intuitionistic metalogic for intuitionistic logic, with results quite different from those of classical metalogic for intuitionistic logic (see McCarty, 2008, for example). Intuitionists start from proof theory, not from model theory. They already know which logical principles they are initially willing to use, and they start reasoning on that basis in the metalanguage as well as the object-language. They seem to have succeeded in adopting intuitionistic logic, internalizing their alternative in a way proponents of many-valued logic may have failed to do.

In 'The Question of Logic', Kripke acknowledges that we can and sometimes do change our views as to which principles are laws of logic. But he insists that in doing so, and in drawing consequences from our new views, we still rely on 'intuitive

reasoning', as he calls it. Clearly, without dispositions to reason in some ways rather than others, we are left logically inert, or floundering about at random. But that does not mean that our dispositions to reason are unmalleable. In particular, it leaves open the possibility that a change in one's views about logic may be part of a more general change in one's ways of thinking, whereby one's dispositions to reason also change. Which patterns of reasoning are intuitive for one may change too. As far as I can see, Kripke's arguments in 'The Question of Logic' do not foreclose that possibility, and are not designed to do so. Perhaps he would not count it as 'adopting a logic' in the sense in which he is arguing that one cannot adopt a new logic. But in that case we need another term for what is happening. I will describe it as 'coming to accept' a new logic. Below, I will say much more about what coming to accept a new logic can involve. However, I will not attempt to sharpen the contrast between 'adopting' and 'accepting' a new logic, because Kripke himself expresses uncertainty about the meaning of 'adopt' (see below), a word which he treats as taken from his opponents' lexicon. Consequently, in holding that someone can come to accept a new logic, I am not sure that I am disagreeing with anything in 'The Question of Logic'; my purpose is not to pick a quarrel with Kripke but to understand change of logic as a special case of theory change.

There is a specific disagreement with Kripke about the case of intuitionistic logic. He writes: 'Intuitionists, I think, are not really proposing to modify the laws of logic with the classical connectives'. Rather, he suggests, they are proposing new connectives with new meanings, not rejecting classical logic as mistaken. He sums up: 'I do not think that this famous case is a case where a different logic was "adopted", whatever that may mean' (Kripke, 2020; note the uncertainty about the meaning of 'adopted'). What Kripke says may apply to some people who call themselves 'intuitionists', but it is hard to square with the writings of people such as Brouwer and Dummett. I will argue that it really is possible (though unwise) to accept intuitionistic logic in a way which involves rejecting classical logic.

'The Question of Logic' also assigns a privileged position to the rule of universal instantiation, since without it statements of logical rules are useless, because we cannot extract their instances. I will argue that logicians can also genuinely differ in what principles of instantiation they allow.

The next section makes a start on the question of accepting a new logic by considering it in the case of alternative modal logics.

## 16.2 Accepting an Interpreted Modal Logic

How does the problem of alternative metalogics, mentioned above, impact on modal logic? In principle, one can try to develop a modal model theory for modal logic, using a metalanguage with modal operators semantically like those in the modal object-language to characterize classes of models for the object-language—just as the standard model theory for first-order non-modal logic uses a metalanguage with first-order quantifiers semantically like those in the object-language to characterize

classes of models for the object-language. In practice, however, that more homophonic approach has made little progress; technically, it is at best desperately cumbersome (Humberstone, 1996 is a brave attempt).

By contrast, the 'possible worlds' approach pioneered by Kripke (1963) has been immensely fruitful, mathematically, philosophically, and in applications to semantics, computer science, and theoretical economics. The metalanguage is the language of set theory, extended by some vocabulary for syntax. It is *non-modal*: it has no modal operators. In that language, one formally defines in set-theoretic terms what a *Kripke model* is. Each such model comes equipped with its own set of what are informally conceived as (possible) worlds, but can in fact be anything. Formally, one defines the truth-value of a formula of the modal object-language at a world in a Kripke model, validity over a class of Kripke models, and so on. By purely non-modal reasoning, one can study which modal formulas are valid over a given class of Kripke models. The standard metalogic of modal logic concerns such mathematical matters, which are independent of any intended interpretation of the modal operators.

Intending the modal operators in a formal language to be interpreted in terms of a given informally specified modality, one may conjecture that the formulas of the language always true on that interpretation are exactly the formulas true in every member of a given mathematically specified class of Kripke models (for more details where metaphysical modality is intended, see Williamson, 2013: pp. 92–118, 130–47). Relying on that conjecture, one may guide one's reasoning about the intended modality by using non-modal reasoning to calculate which formulas are true in every model in the specified class. But the conjecture needs to be tested. In doing so, one tests the choice of models against informal modal judgments at least as much as one tests informal modal judgments against the choice of models. The chosen class may just be used to *systematize* some initial modal judgments, and to work out which other modal judgments follow from, or are consistent with, that systematization. But those services may also serve a more critical purpose, by making some of the initial modal judgments look anomalous, or revealing them to be in tension with each other.

Interpreted modal logics are in effect austere structural *theories* about the intended modality, such as metaphysical or practical possibility. Arguably, the appropriate methodology for deciding between rival theories of such a subject matter is *abductive*, like that for deciding between rival theories in other sciences (Williamson, 2013, pp. 423–9). The initial evidence comes from our pre-theoretic knowledge of the intended modality in particular cases. For example, we have pre-theoretic knowledge of practical possibility, often expressed by 'can'. We know that some people can run a mile in under four minutes; of course, they can also run a mile in over four minutes. We know that you do only what you can. We know that if you can run and cannot fly, you can run without flying. And so on. A modal logic proposed for a given modality is expected to systematize such already known truths. It must be consistent with them; the more it unifies them by bringing them under strong and simple generalizations, the better.

As in other sciences, answering theoretical questions may involve various kinds of indirect strategy. For instance, a central dispute in modal metaphysics is between *necessitism* and *contingentism*. Necessitism is expressed by the formula

$\Box\forall x\Box\exists y(x = y)$, where $\Box$ expresses metaphysical necessity and the quantifiers range over absolutely everything; it says that necessarily everything is necessarily something. Contingentism is just the negation of necessitism. They are rival answers to the question 'Is there contingency in what individuals there are?' Each of them is a broadly logical claim: it is expressed without non-logical constants (counting $\Box$ as a broadly logical constant). Under reasonable assumptions, the validity of necessitism is equivalent to the validity of reasoning by interchanging $\Diamond\exists v$ and $\exists v\Diamond$, or the Barcan formula and its converse (see Barcan, 1946 and Kripke, 1963; such interchanges of $\Diamond$ and $\exists$ were in effect already endorsed by Ibn-Sina = Avicenna, see Williamson, 2013: p. 45). Necessitism corresponds to the class of Kripke models for quantified modal logic in which the domain of individuals is constant across worlds; contingentism corresponds to the broader class of models with no such restriction (Kripke, 1963). Arguably, our pre-theoretical knowledge is insufficient to settle the issue between necessitism and contingentism. A more systematic, theoretical case is needed for one side or the other; it may take hundreds of pages to build. One's commitments may shift as new considerations come into view. I first encountered the issue by reading Kripke (1963). At first, I inclined to contingentism because I found the apparent common sense counterexamples to necessitism metaphysically compelling. Later reflection persuaded me that necessitism can do justice to such examples, and I came to prefer it on abductive grounds: necessitism beats contingentism in its combination of simplicity and strength (Williamson, 2013). That certainly *felt* like adopting an interpreted modal logic. But perhaps I just accepted an interpreted modal system, without adopting it. What matters for present purposes is that I *did* come to accept a modal system which I had previously rejected (under the same interpretation of the logical constants, including $\Box$), and to reason by underived modal principles which I had previously refused to reason by.

We can see accepting an interpreted modal system as intermediate between accepting a theory in natural science and accepting a mathematical theory such as Zermelo-Fraenkel set theory with the Axiom of Choice. In interpreted modal logic, like mathematics and unlike natural science, evidence from experiment, observation, and measurement normally plays no special role in accepting a theory. On the other hand, in interpreted modal logic, like natural science and unlike mathematics, disputes between rival theories are both usual and central to the development of the field. Obviously, as those contrasts underline, the methodology of theory acceptance is far from uniform across the sciences, including both the natural sciences and mathematics. Nevertheless, interpreted modal logic does not look like an *outlier* with respect to theory acceptance.

Of course, modal logic might be an atypical branch of logic. Why might 'core' first-order non-modal logic be different?

One reason is that first-order non-modal logic seems to be needed as part of the background logic for just about any theory. Without it, how can we extract consequences from the theory's postulates? Although its apparent ubiquity does not by itself prove that one cannot change one's system of first-order non-modal logic, it might make accepting such a system somehow special. Section 16.4 bears on that issue.

Another reason why modal logic might be atypical is that accepting an interpreted modal system is usually a matter of assessing various potential *axioms*, not rules of inference. The background non-modal logic is normally axiomatized with at least one rule of inference (such as modus ponens), but the latter is rarely at stake in the choice of modal system: the modal principles are the issue. Although most modal systems are standardly axiomatized with the rule of necessitation (if $A$ is a theorem, so is $\Box A$), Kripke has shown how it can be eliminated for normal modal systems, by replacing each axiom $A$ with the infinite family of axioms $A, \Box A, \Box\Box A, \ldots$(Kripke, 1963). By contrast, rules of inference are often central to the choice of a first-order non-modal system. The focus of Kripke's account of the adoption problem is on rules of non-modal inference such as universal instantiation and modus ponens, which one applies to extract the required consequences from the formal specification of the system to be adopted (Kripke, 2020). Psychologically, a rule of inference corresponds not to a belief or even an infinite family of beliefs but to a pattern of *transitions* between beliefs (Carroll, 1895, which Kripke, 2020 cites; see Besson, 2019 for some relevant discussion). The difference between believing and making transitions between beliefs might have repercussions for the comparison between accepting a system of logic and accepting a theory in other sciences.

Sections 16.3–16.5 will argue that theory acceptance in logic is *not* exceptional, in relation to theory acceptance elsewhere. Before that, some preliminary clarifications will be useful.

First, in speaking of 'theories' in logic, we will always mean *interpreted* logical theories, where each logical constant (including any modal operator) has an intended interpretation. Uninterpreted formal systems can be studied mathematically, but they are not normally candidates for acceptance or rejection.

Second, we must be careful not to confuse *logical* theories with *metalogical* theories. Even though we count all metalogical theories as logical (in a broad sense), we must not count all logical theories as metalogical. For instance, one theorem of an orthodox logical theory of identity is $\forall x\ x = x$, which just says that everything is identical with itself (for definiteness, we may read the quantifier as unrestricted). That is *not* a metalogical statement; it says no more *about* logic, formulas, proofs, or models than it says about cabbages and kings (by implication, that they are self-identical). In particular, it must not be confused with a metalogical statement such as $\vdash \forall x\ x = x$ (which says that the formula $\forall x\ x = x$ has a proof of the relevant formal kind), or $\models \forall x\ x = x$ (which says that the formula $\forall x\ x = x$ is true in every model of the relevant formal kind), or the less formal '$\forall x\ x = x$ is a logical truth'. Although $\forall x\ x = x$ is a logical truth, it does not say of itself that it is such; its meaning is compositionally derived from the meanings of its constituents; neither the universal quantifier, nor the identity sign, nor the variable '$x$' imports anything distinctively metalogical into the meaning of formulas in which they occur. Correspondingly, to *think* 'Everything is identical with itself' is not to think 'It is a logical truth that everything is identical with itself', otherwise an infinite regress would threaten: to think 'It is a logical truth that everything is identical with itself' would be to think 'It is a metalogical truth that it is a logical truth that everything is identical with itself', and so on.

Similarly, although perhaps a little less obviously, to think *C* having inferred it from *A* is not to think '*C* is a logical consequence of *A*' (even if *C* is in fact a logical consequence of *A*). In general, a theory can enable one to mount good arguments without being a theory *about* which arguments are good. For example, a simple physical theory may enable one to infer 'This will not float in water' from 'This is a rock' (with the reference of 'this' held fixed). That does not show that the physical theory is about such argumentative connections; it may be just about things like water, rocks, and floating.

In what follows, logical theories and inferences based on them are assumed non-metalogical unless otherwise specified. Such logical theories are *not* theories of validity, which are metalogical (contrast Hjortland, 2017). Rather they are theories of what the logical constants refer to: negation, conjunction, disjunction, conditionality, universality, existence, identity, various kinds of possibility and necessity, and so on. Many of these can be understood as functions from propositions to propositions: for example, possibility maps the proposition that Leibniz was a polymath to the proposition that it is possible that Leibniz was a polymath. Others map properties to propositions: for example, universality maps the property of changing to the proposition that everything changes. Yet others are relations, which in effect map individuals to propositions: for example, identity maps Socrates and Plato to the proposition that Socrates is Plato. A more accurate but cumbersome statement of all this would be in higher-order terms, dropping the nominalizations in favour of the logical constants themselves. Good theories of such matters tell us something general about the very abstract structure of reality.

Granted, metalogic is a branch of logic, and tells us much about logical validity. That is a topic of great depth but much narrower interest than the more basic parts of logic just described. Theories of virtually all kinds use the standard logical constants, and so are sensitive to the behaviour of the functions and relations to which they refer, even when it is not the topic under discussion. Of course, once we start consciously monitoring our own reasoning, we have to reflect on questions of validity. Even then, however, we need not distinguish logical from non-logical validity. For most purposes, all we need is the more general distinction between deductively good and bad reasoning. For instance, the two forms of reasoning '*t* is red, so *t* is red or blue' and '*t* is red, so *t* is coloured' are both deductively good (necessarily truth-preserving) for almost all purposes, but only the former is *logically* valid, because it depends just on the meaning of the logical constant 'or', whereas the latter depends on the meanings of the non-logical constants 'red' and 'coloured'. In that sense, ordinary reasoners can easily get by with no idea of *logical* validity; they cannot easily get by with no idea of negation, conjunction, identity, and the like.

Ordinary speakers do not understand technical jargon such as 'logical consequence' and 'logical constant'. They do understand logical constants such as 'not', 'and', 'or', 'if', 'is', 'all', 'some', 'can', and 'must' (though the status of the last two as logical constants is disputed). Humans have evolved a moderately reliable capacity to assess simple statements about their environment made using such logical constants in combination with non-logical terms as true or false. They have also evolved a moderately reliable capacity, by pattern recognition, to assess simple inferences between

such statements as good or bad reasoning (non-accidentally truth-preserving or the opposite). Such pattern recognition makes a natural starting-point for inquiries which may eventually mature into systematic logic, with the positively assessed statements as an evidence base, to check for counterexamples. Similarly, humans have evolved a moderately reliable capacity to assess simple statements about their environment made using natural kind terms and spatio-temporal vocabulary in combination with other terms as true or false. They have also evolved a moderately reliable capacity, by pattern recognition, to assess simple inferences between such statements as good or bad reasoning (non-accidentally truth-preserving or the opposite). Such pattern recognition makes a natural starting-point for inquiries which may eventually mature into systematic natural science, with the positively assessed statements as an evidence base, to check for counterexamples. This picture suggests the main idea of the next section.

## 16.3 Anti-exceptionalism About Logic

The vague hypothesis inspiring this paper is anti-exceptionalism about logic: amongst the sciences, logic is no outlier; it differs from the rest only as much as they differ from each other.

The classic statement of a version of anti-exceptionalism about logic is of course Quine's, in 'Two Dogmas of Empiricism' (Quine, 1951), to which 'The Question of Logic' is in part a response, turning against Quine a style of argument which Quine had deployed against Carnap's conventionalism about logic (Quine, 1936). The present version of anti-exceptionalism differs from Quine's. In particular, his 1951 article contains a strong residue of empiricism; his anti-exceptionalism about logic is articulated in the final section, entitled 'Empiricism without the Dogmas'. Sensory experience is still described as the boundary condition for total science. That experience has any such unique privilege is not a truism, unless 'experience' is just another word for *learning*. On an alternative view, the evidence we have to go on, in both science and daily life, is simply whatever we happen to know, however we happen to know it (Williamson, 2000, pp. 184–208). The present version of anti-exceptionalism is not a form of empiricism about logic, since it ascribes no special privilege to experience, even if in practice most human knowledge is inseparable from sensory experience. However, distancing this version of anti-exceptionalism about logic from empiricism does not by itself immunize it against Kripke's critique, which engages with a different aspect of Quine's version.

My aim is to sketch a picture of theory acceptance on which we can and do accept theories in our logical reasoning, just as we can and do accept theories in our physical reasoning, and we can and sometimes do *change* which theories we accept in our logical reasoning, just as we can and sometimes do change which theories we accept in our physical reasoning. Now if we follow the standard logical positivist identification of a theory with a set of sentences (as Quine tended to do), the dynamic aspect of reasoning seems to be left external to theories, so that mere changes of

theory would leave reasoning itself untouched. That applies as much to theories in logic as to theories elsewhere. To deal with this difficulty, we must understand that, on a more natural, less formalistic conception of theories, they themselves often have a dynamic aspect: they include patterns of reasoning, not just sentences describing such patterns.

In developing the analogy between logic and physics, it is useful to recall the standard distinction between *scientific physics* and *folk physics* (also known as 'naïve physics' and 'intuitive physics'; see for example Smith & Casati, 1994; Baillargeon, 2004; Spelke & Kinzler, 2007; Gelman & Noles, 2011, and Kubrich et al., 2017). Folk physics comprises the instinctive, pre-reflective forms of physical reasoning on which humans rely, independently of their education. It is more or less innate and universal to the species. For example, we typically use folk physics in automatically forming expectations about the movement of inanimate objects in our environment. By contrast, scientific physics comprises non-instinctive, reflective forms of physical reasoning which human may learn, depending on their education. It is neither innate nor universal to the species. The distinction is of course not perfectly precise, and it is complicated by various factors: for example, learnt scientific forms of physical reasoning can gradually become habitual. Through reflection, scientific physics can sometimes override and correct folk physics. The process of accepting a new theory in physics pertains to scientific rather than folk physics, though once the new theory is accepted, thinking in accordance with it may of course become habitual.

We can draw an analogous distinction between *scientific logic* and *folk logic* (naïve logic, intuitive logic). Folk logic comprises the instinctive, pre-reflective forms of logical reasoning on which humans rely, independently of their education. It is more or less innate and universal to the species. For example, we typically use folk logic in automatically making very simple deductions from what we know. By contrast, scientific logic comprises non-instinctive, reflective forms of logical reasoning which human may learn, depending on their education. It is neither innate nor universal to the species. The distinction is of course not perfectly precise, and it is complicated by various factors: for example, learnt scientific forms of logical reasoning can gradually become habitual. Through reflection, scientific logic can sometimes override and correct folk logic. The process of accepting a new theory in logic pertains to scientific rather than folk logic, though once the new theory is accepted, thinking in accordance with it may of course become habitual.

An example where scientific logic may have overridden folk logic concerns the form of inference from 'Every F is G' to 'Some F is G', standardly rejected as invalid by both modern logic and modern semantic theories of natural language. Kripke (2020) interprets the episode as one where Aristotelian proto-scientific logical theory erred, by neglecting the case of an empty subject term, and was corrected by intuitive reflection. However, that interpretation is hard to square with the fact that the consensus was achieved only in modern logic. Medieval logicians quite capable of intuitive reflection sided with Aristotle, not by neglecting the empty case but by denying the truth of the universal generalization in that case. Nor is the problem resolved by the claim that we *presuppose* the non-emptiness of the subject term, for pre-theoretically both the inference from 'There is no F' to 'There is no F which

is not G' and the inference from there to 'Every F is G' seem good, while the original inference still seems good too. More likely, folk logic relies on quick and easy *heuristics*, efficient but less than perfectly reliable forms of reasoning. Their status as heuristics need not be transparent to the naïve user; when they lead one into inconsistency, one may have no pre-theoretic way of locating the source of the error. Similar problems arise for conditionals. Just as folk logic may erroneously regard 'Every F is a G' and 'No F is a G' as contraries, so it may erroneously regard 'If A, C' and 'If A, not C' as contraries too. A more abstract, systematic, and theoretical approach may be needed to resolve the issue (see Williamson, 2020 for further development of these themes). One of Aristotle's skills seems to have been in articulating good approximations to folk theories, in logic as in physics.

By 'accepting' a theory (in physics, logic, or anywhere else) I mean more than just accepting it for the sake of argument or some other limited purpose. In particular, I mean more than just accepting the theory as 'empirically adequate', which is in any case a rather obscure status, given the deep unclarity of the distinction between 'empirical' and 'non-empirical' knowledge (compare van Fraassen, 1980 and Boghossian and Williamson, 2020). I mean something like believing it, relying on it, being disposed to apply it and act on the results, much more generally. That is all-out acceptance. Of course, one can also take various attitudes of qualified or partial acceptance, to which the arguments below will generalize in suitable ways.

Why not simply speak of 'believing' a theory, rather than 'accepting' it? Recall the distinction between axioms and rules of inference. It is fine to speak of 'accepting' an axiom and of 'believing' the axiom. It is also fine to speak of 'accepting' a rule of inference, but it is *not* fine to speak of 'believing' a rule of inference; it does not seem to be the right kind of thing to be believed. Of course, one can believe a rule of inference *to be valid*, but that is a metalogical belief. By contrast, accepting a rule of inference—like accepting an axiom—is not metalogical. For example, one accepts the rule of modus ponens by relying on it in one's reasoning; in doing so, one thinks the premises and conclusion, but normally one does not think about validity: one might think about it if one started having reflective *doubts* about modus ponens. Thus 'accept' is a better word than 'believe' for the non-metalogical relation one has to both axioms and rules of inference.

Invoking the distinction between 'axioms' and 'rules of inference' may suggest that we are talking about logic, but it makes sense for theories in physics too. Of course, the axioms are not required to be self-evident; they may simply be *postulates*, not derived from anything else. The category of rules of inference seems especially apt for folk physics. For example, if you want a device which takes as input the presently seen position, velocity, and direction of an object and rapidly delivers as output its predicted position, velocity, and direction one second into the future, a rule of inference is exactly what you need. Obviously, the agent is not aware of any formula by which the prediction is made; the agent does not need the ability to articulate the rule.

Naturally, scientific physics is more articulate than folk physics. Even so, Thomas Kuhn's influential account in *The Structure of Scientific Revolutions* of what it is to accept a scientific paradigm suggests that much of it involves recognitional capacities

for types of problem and ways of solving them which the scientists may be unable to articulate properly in words (Kuhn, 1962). The category of (loose) rules of inference may be needed there too.

Needless to say, this appeal to Kuhnian descriptions involves no commitment to the gratuitously relativist and constructivist remarks with which Kuhn tended to betray his lack of philosophical training. Nor is it to endorse the vulgar misreading of his work as showing (or aiming to show) that science is as irrational as everything else; rather, he offered a more sophisticated account of how individual irrationality can contribute to a collectively rational process—for example, when the obstinate loyalty of scientists to a paradigm in which they are invested, despite all the anomalies they cannot explain, helps ensure that its intellectual resources are thoroughly explored.

On this account of theory acceptance, it is seriously misleading to picture it as like writing a sentence or list of sentences in one's 'belief box'. Still, it is helpful to keep in mind the connection with belief. In particular, the metaphor of adoption suggests a *voluntary* process (as does the metaphor of writing in one's belief box), whereas belief is notoriously involuntary. One cannot just switch on the belief that one is ten metres tall, or switch off the belief that one is much less than ten metres tall, however much money one is offered to do so. Indirectly, one can sometimes manipulate oneself into or out of a belief, but that is not the same thing. Since acceptance is a generalization of belief, the involuntariness of belief is likely to generalize to the involuntariness of acceptance. The evolutionary pressures towards involuntariness are the same in the two cases: we act on what we believe or accept, so it is dangerous to leave such matters to individual caprice. In more epistemological terms, both beliefs and acceptances are normally products of processes whose function is to produce knowledge, whereas voluntary beliefs and acceptances would normally not be knowledge. Thus, if adoption is voluntary, the idea of adopting a logical theory will be suspect.

Admittedly, it is not clear that belief is *always* involuntary. For example, imagine a member of the jury in a murder trial. From early in the trial, there is strong evidence that the accused is guilty, enough by ordinary standards to justify a bystander in believing in his guilt, but perhaps not quite enough by legal standards to secure conviction. The juror is strongly inclined to believe him guilty. However, she has a keen sense of responsibility, and wants to hear the remaining evidence and arguments with an open mind. She may therefore decide to hold off believing that he is guilty, deliberately suspending judgment until she has heard both sides of the case in full. Alternatively, she may decide that doing so would be over-scrupulous, and just go ahead anyway and believe him guilty. In such a case, she may genuinely be deciding whether to believe that he is guilty. There is nothing obviously psychologically impossible about the case (see Ginet, 2001; van Fraassen, 2002; Weatherson, 2008; Rott, 2017; and Williamson, 2024 for further discussion and references).

One can construct analogues of the juror case for believing a theoretical proposition or accepting a theory, in physics or logic, where the evidence is strong but perhaps not quite strong enough. There may be just enough wiggle room for voluntariness. For *partial* acceptance, for a limited purpose, the scope for voluntariness may be quite large.

However, even if such cases are possible for all-out acceptance, they are somewhat marginal. In most cases, whether one fully believes a proposition or accepts a theory is not a matter of will—which is not to deny that it is a matter of rationality. Thus, if adopting a theory in logic is supposed to be a matter of will, it faces an obvious problem. Likewise, if adopting a theory in *physics* is supposed to be a matter of will, it faces the same obvious problem. There is no problem for anti-exceptionalism about logic here. We will finesse this issue by not assuming that theory acceptance is a matter of will.

With these clarifications in place, the next section aims to get a little more realistic about theories and what it is to accept them, in order to identify underlying similarities between theories in logic and theories elsewhere, which the disciplinary stereotypes may conceal.

## 16.4 Theories and Dispositions

The talk in previous sections of 'axioms' and 'rules of inference' suggests something much more precise and perspicuously organized than most extant theories aspire to be. At first, this point may seem to tell *against* anti-exceptionalism about logic. However, properly understood, it enables us to see deeper similarities between accepting a theory in logic and accepting a theory elsewhere.

Most theories have never been axiomatized, and may not be susceptible to axiomatization in any useful way. Many theories hardly fit the standard metalogical definition of a 'theory' as a set of sentences closed under logical consequence. Consider Darwin's theory of evolution. Although attempts have been made to axiomatize it, they have no special status in biology; probably most biologists have never heard of them. When one asks what the theory says, one may be told something like 'Species evolve by natural selection'. That sounds more like a *generic* generalization than a universal one: its truth does not entail the truth of every instance ('Ducks lay eggs' is a standard example). The theory of evolution would not be refuted by a scientist artificially creating a species and then destroying it before it had the chance to evolve. As a generic generalization, it does not deductively entail very much. Yet, by reasonable standards, it is an extremely powerful explanatory theory. Where does this power comes from? A natural suggestion is that it comes from the cognitive *dispositions* involved in fully accepting it, such as dispositions to seek, develop, and accept specific explanations by natural selection—explanations in the spirit of the theory but not entailed by it, even when it is combined with background knowledge. Nor should we expect the theory to provide a precisely articulated *rule* to tell us which such explanations to accept.

Faced with such a theory, whose propositional content is so elusive, some philosophers might describe what it offers us as 'understanding' rather than 'knowledge'. However, that is a false contrast, based on too narrow a conception of knowledge. One cannot understand why or how things happen without *knowing* why or how they happen (see Sliwa, 2015).

Obviously, the theory of evolution is a very rich and complex case, to which only an accomplished philosopher of biology could do proper justice. It may be helpful to consider a vastly simpler theory, which is widely accepted too, though (unlike Darwin's) not for good reasons. The *conspiracy theory of history* says something like this: 'Historical events result from conspiracies'. That too is a generic generalization, not a universal one: its truth does not entail the truth of every instance. Proponents of the conspiracy theory can consistently admit that *not all* historical events result from conspiracies. Yet, whenever a particular historical event comes under discussion, they will typically explain it as the result of a conspiracy. Although the generic generalization 'Historical events result from conspiracies' does not deductively entail very much, its explanatory power comes from the cognitive dispositions involved in fully accepting it, such as dispositions to seek, develop, and accept specific explanations by conspiracy—explanations in the spirit of the theory but not entailed by it, even when it is combined with background knowledge. Nor should we expect the theory to provide a precisely articulated *rule* to tell us which such explanations to accept.

Imagine someone who assents to the generic slogan 'Historical events result from conspiracies' if the question arises but, when it comes to cases, is not at all disposed to prefer explanations of particular historical events by conspiracy to other explanations of them. Their acceptance of the conspiracy theory of history seems at best half-hearted. They may just be paying it lip service. Conversely, imagine someone who dissents from the generic slogan if the question arises but, when it comes to cases, is strongly disposed to prefer explanations of particular historical events by conspiracy to other explanations of them. Their rejection of the conspiracy theory of history seems at best half-hearted. They may be paying it all but lip service. In such cases, how far one accepts the generic theory depends as much on one's dispositions when it comes to cases as on one's reaction to the generic slogan itself.

The foil usually offered to the conspiracy theory of history is not a grandiose theory of history like those of Marx and Tolstoy but rather the plain *cock-up theory of history*, for which I have a sneaking sympathy. It says something like this: 'Historical events result from cock-ups [incompetence, confusion, mistakes, misunderstandings, etc.]'. What it is to accept it can be characterized in terms exactly analogous to those just used to describe what it is to accept the conspiracy theory.

Switching between the conspiracy and cock-up theories of history may be rare, but is clearly possible, in both directions. It could take the form of a gradual change in one's disposition to prefer one type of explanation over the other when it comes to cases, with one's switch from one generic slogan to the other just an after-effect.

How different is theory acceptance in logic and mathematics, where theoretical generalizations are universal rather than generic? The difference is less stark than one might have expected.

Many working mathematicians have no special concern for foundational issues. They rely on set theory as the standard framework for mathematics but are not interested in it for its own sake. Perhaps they have briefly encountered Russell's paradox as an exotic curiosity and are vaguely aware that set theorists have ways of dealing with it, but they may never have seen the axioms of a formal set theory such as ZFC, and be incurious about what the accepted theory is. In effect, such mathematicians

have a defeasible disposition to accept instances of the Naïve Comprehension principle, to help themselves to a set demarcated by a predicate when they need it for their work, while subliminally aware that doing so is not a universally valid rule, since it generates contradictions by Russell's paradox. They may reject 'funny-looking' instances of Naïve Comprehension; such instances do not crop up naturally in their work. They have no formal rule to determine whether an instance does look funny, any more than we consult a rulebook to determine whether to laugh at a joke. They implicitly rely on Naïve Comprehension as a generic generalization rather than a universal one. Their dispositions correspond to no standard restricted Comprehension principle, in part because they work with a much wider vocabulary than a formal language for set theory, for example when they apply mathematics to the natural or social world.

Years, arguably decades, passed between the emergence of the set-theoretic paradoxes and the solidification of a consensus amongst mathematicians that the methods of Russell or Zermelo were adequate for containing them. In that period, a similar attitude to principles of set theory was rational even for mathematicians concerned about foundational issues. For example, Whitehead expressed such an attitude to the theory of classes in applying mathematics to the material world, claiming that his applications would be robust under different resolutions of the paradoxes (Whitehead, 1906, p. 470). For logicists, taking that attitude to mathematics is tantamount to taking it to logic.

We remain in a similar position with respect to the logic of predicates for truth and falsity, faced with the Liar and other semantic paradoxes, despite much progress in the tradition of Tarski (1935) and Kripke (1975). There is still no consensus as to exactly what approach to follow. We use instances of disquotational principles when we need them in practice, while aware that the unrestricted application of those principles will lead to contradictions. Typically, we do so with no particular formal theory of truth in mind. Rather, we treat the disquotational principles more like generic generalizations. We have a defeasible disposition to accept their instances, which we overrule for funny-looking instances. As usual, we have no formal rule to determine whether an instance does look funny: how would such a rule handle the indefinite semantic diversity of sentences in English or any other natural language to which we want to apply disquotation?

This need for genericity in logic arises even for first-order reasoning in natural languages. Consider inferences by the rule of disjunction introduction, to a disjunction from one of its disjuncts. We have a defeasible disposition to accept inferences of that form. However, we may override our disposition to infer from 'P' to 'P or Q' when 'P' is fine but 'Q' looks nonsensical, paradoxical or otherwise funny, perhaps also when it contains a racial or other slur (read the quotation marks as Quinean corner quotes). Once again, we have no formal rule for detecting such exceptions.

Similar issues arise for the rule of universal instantiation, which plays a central role in Kripke's argument about adoption. We have a defeasible disposition to make inferences of the form of universal instantiation, say from 'Everything is F' to 't is F', where 't' is a singular term. However, we may override that disposition when 't'

is funny-looking, somehow suspect. For instance, I am happy to accept 'Everything is self-identical', but reluctant to infer 'Satan is self-identical' (I do not believe in devils).

Of course, such examples are the staple diet of proponents of 'free logic', on which one is not free to apply universal instantiation. For them, the valid move in the vicinity is from 'Everything is F' and the extra premise 'Something is t' to 't is F'; since I do not accept 'Something is Satan', no wonder I am reluctant to infer 'Satan is self-identical'. However, free logic does not remove the need for genericity. For, in order to vindicate the default status of unrestricted universal instantiation (without which free logic would be hopelessly restrictive), free logic must assign default status to 'Something is t'. We will override the disposition to accept 'Something is t' in the same cases, when 't' is suspect. Clearly, we have no formal rule for detecting doubt as to whether 't' refers.

The case of free logic shows that the classical rule of universal instantiation is not an utterly non-negotiable feature of our 'intuitive' reasoning. We could make do by reasoning in a free logic, with a strong enough but still defeasible disposition to accept the extra premises. For many purposes, we could even make do by reasoning with 'universal' generalizations as if they were merely generic generalizations, whose truth is compatible with a few exceptions, according to the principles of a logic in which 't is F' follows defeasibly from 'Everything is F' and background assumptions, unless the latter include defeaters. Such a logic is non-monotonic and so non-deductive, in the sense that adding extra premises can make a valid inference into an invalid one. Predictably, I do not *endorse* any such way of weakening classical universal instantiation. For scientific purposes, the 'counterexamples' to classical universal instantiation involve defects in the language, such as reference failure, which are better handled by not reasoning with the defective terms than by complicating the basic principles of reasoning merely to allow oneself the dubious advantage of reasoning with defective terms. Moreover, the standard exception-intolerant non-generic reading of the universal quantifier is exactly what one needs for many scientific purposes, not least in mathematics. Thus the proposed revisions of the rule of universal instantiation are bad science. But bad science is not *impossible*. Free logicians may have genuinely internalized the principles of their preferred free logic, and come to 'reason intuitively' according to them, so not to rely on the classical rule of universal instantiation.

All these examples indicate that accepting principles of mathematics or logic outside a formal language often involves messily defeasible dispositions not utterly unlike those involved in accepting theories elsewhere. Such dispositions are not normally immutable. We can consider some cases of changing one's dispositions to reason in various ways in mathematics and logic.

Here is a toy example. A child is learning elementary mathematics. On the basis of recognizing a pattern, the child becomes disposed to accept any instance of '$n/n = 1$'. On the basis of recognizing another pattern, the child also becomes disposed to accept any instance of '$0/n = 0$'. One day, the child accepts '$0/0 = 1$' as a result of the former disposition and '$0/0 = 0$' as a result of the latter disposition, which prompts the alarming thought '$0 = 0/0 = 1$'. On reflection, the child restricts both

dispositions. At first, the restriction may take the form of retaining the original dispositions while adding a secondary disposition to inhibit them for $n = 0$. Later, the restriction becomes habitual: the original incorrect dispositions may disappear, being replaced by correctly restricted dispositions; the child is no longer even tempted to accept '$0/0 = 1$' or '$0/0 = 0$'. That is a case of *correcting* one's dispositions. But if it is possible to correct one's dispositions, it is also possible to have the *illusion* of correcting one's dispositions, while in fact replacing correct dispositions by incorrect ones.

At a more advanced stage, the child has learnt to add, subtract, multiply, and divide correctly. He can do simple algebraic manipulations on the basis of pattern recognition; for instance, from '$k + m = k + n$' he infers '$m = n$'. He can distinguish natural numbers from other numbers, also on the basis of pattern recognition. However, he has not learnt any proof system for arithmetic. Moreover, he has no disposition to reason by mathematical induction. When the teacher first explains the principle of mathematical induction to him, it does not immediately strike him as valid. He has no particular objection to it; he just does not find it compelling (he is not terribly intelligent). Nevertheless, he accepts mathematical induction on the teacher's authority, and learns to use it himself. Reasoning by mathematical induction gradually becomes habitual to him, and he feels confident in such reasoning. He has acquired a disposition to reason by mathematical induction. That is a case of gaining a disposition. But a disposition so gained may also be lost. If an eccentric new mathematics teacher tells the class that recent developments in mathematics have discredited the principle of mathematical induction, the pupil will shrug his shoulders and give up such reasoning.

Now imagine an unwise philosopher, Phil, who gets it into his head that verification is *constructive*: to verify a disjunction you must verify a disjunct, and to verify an existential generalization you must verify an instance. As he recognizes, this makes trouble for the law of excluded middle. For when we have no verification of 'S' and no verification of 'not S', we have no verification of 'S or not S'. With the courage of his convictions, Phil refuses to accept excluded middle; he starts training himself not to assume it in his reasoning, and not to assume principles from which he can derive it. He hears about intuitionistic logic, which sounds like just what he needs. He starts imitating intuitionistic reasoning. Phil's thinking is much cruder than that of philosophers like Michael Dummett, attracted by intuitionistic logic as an alternative to classical logic; unlike them, Phil has no category of indirect verification for cases where one possesses a procedure guaranteed to generate a verification of one disjunct or the other, but has not yet applied it to the case at hand. But it is still *psychologically possible* to think in Phil's crude way; anyway, even acknowledging the category of indirect verification is insufficient to validate the law of excluded middle, since one may have no procedure of the kind needed for indirectly verifying 'S or not S'. From time to time I meet self-identified intuitionists whose philosophical thinking seems little better than Phil's.

Phil does not change his dispositions at a stroke. He works on them. His dislike of excluded middle enables him to apply a kind of aversion therapy to himself. There may be an unconscious level of automatic reasoning which his self-improvement

exercises cannot reach, just as theoretical physicists retain some level of instinctive proto-Aristotelian folk physics, but in both cases one can habituate oneself to another form of reflective thinking, which can override one's more primitive instincts. In some sense Phil may not *adopt* intuitionistic logic, but he surely comes to accept it.

Of course, merely losing some of one's past beliefs need not involve disagreeing with one's past self. Since every theorem of intuitionistic propositional logic is also a theorem of classical propositional logic, and classical propositional logic is consistent, no formula $\alpha$ is a theorem of one logic while $\neg\alpha$ is a theorem of the other. Indeed, the negation $\neg(\alpha \vee \neg\alpha)$ of any instance of excluded middle is as inconsistent intuitionistically as it is classically. Thus the intuitionist cannot consistently deny any instance of excluded middle. For plain propositional logic, any disagreement is at another level, either metalogical—the classical logician says that $\alpha \vee \neg\alpha$ is a logical truth while the intuitionistic logician says that it is not—or epistemic—the classical logician says that $\alpha \vee \neg\alpha$ is known while the intuitionistic logician says that it is not. However, once quantifiers are added to the object-language, disagreement can arise at that level too. For intuitionistic mathematics has theorems of the form $\neg\forall x(Fx \vee \neg Fx)$, while of course the classical mathematician asserts $\forall x(Fx \vee \neg Fx)$. The disagreement becomes even starker once quantification into sentence position is permitted in the object-language. The classical logician asserts the general law of excluded middle as a universal generalization, $\forall P(P \vee \neg P)$, while the intuitionistic logician asserts the negation of the law, $\neg\forall P(P \vee \neg P)$. Thus the disagreement can be made explicit in a suitably expressive object-language, without ascent to a metalanguage.

More generally, *differences* between logics with the same vocabulary can be expressed as explicit metalogical *disagreements*, and often manifest non-metalogical disagreements expressible in a suitably extended object-language.

Excluded middle is not exceptional in its contentiousness. Elsewhere, I have argued—with particular attention to modus ponens and the rules for conjunction—that any principle of logic can be rejected by someone who understands the relevant logical words by normal standards (Williamson, 2007; Boghossian and Williamson, 2020). Given its role in the adoption problem, the rule of universal instantiation is of special interest here. It was briefly discussed above; as noted, self-identified free logicians reject it in its usual form. In the reverse direction, one might wonder how someone who did not already implicitly accept universal instantiation could ever come to accept it. But a free logician who becomes disillusioned with all the extra complications of free logic might decide that they are not worth the trouble, and revert to the standard logic of quantification. That might be quite easy: she still remembers how she used to reason, in the good old days before she became a free logician, so she can *imitate* her earlier self. In general, the human capacity for imitation provides a mechanism for taking on different ways of reasoning with no need to derive them from verbal instructions.

By imitation, one can come to accept new patterns of reasoning and reject old patterns. Similarly, one can also come to *simulate* accepting new patterns of reasoning and rejecting old ones, perhaps by actually accepting the new and rejecting the old in very limited contexts. Thus intuitionistic logicians can simulate classical reasoning

and classical logicians can simulate intuitionistic reasoning. Each side imagines how things look, how far they fall into shape, from the other's perspective. Such simulations matter, because they enable us to make informed choices between different ways of reasoning: changing sides is not just taking a leap into the dark. By normal human standards, it is a rational process. In that respect, it is comparable to the kind of process by which scientists sometimes decide to switch from an old theory to a new one in a scientific revolution. In neither case should we expect all-purpose formal rules to tell us which way to jump. Those who demand much more for theory change to be rational may find that the history of science fails to meet their requirements.

Of course, in the case of intuitionistic logic, I regard switching to it from classical logic as a disastrous error. But that does not mean that I cannot imagine how a sane person can regard it as seeing the light. In other cases—such as necessitist higher-order modal logic—I do regard the switch as more like seeing the light.

## 16.5 Denying the Doctrine and Changing the Subject

So far, 'alternative logics' have been treated as genuine rivals to classical logic. Famously, Quine rejected that assumption in his later work. He suggested that would-be alternative logicians can always be more charitably reinterpreted as changing the subject rather than denying the doctrine (Quine, 1970, pp. 82–3; Morton, 1973 is a detailed critique of his argument). For we need not assume that they mean what we mean by logical words such as 'not', 'if', 'and', 'or', 'is', 'every', and 'some' in a natural language, or their analogues in a formal language.

The 'logical pluralism' of Beall and Restall (2006) also involves a charge of equivocation, but it is much more limited in scope than Quine's, and in particular focusses on metalogical vocabulary (such as 'valid') rather than logical vocabulary in the object-language itself.

Quine's charge of equivocation gains some plausibility from 'alternative' logicians' tendency to provide highly non-homophonic semantic theories, to explain why their systems accept some principles and reject others. If those semantic theories characterize meanings for the logical words different from the meanings characterized by the corresponding homophonic semantics, which seem to be the usual meanings, then they characterize *un*usual meanings for the logical words. Kripke discusses the equivocation hypothesis for both intuitionistic and quantum logic, briefly but sympathetically (Kripke, 2020).

However, other aspects of 'alternative' logicians' practice tell the other way. They typically accuse classical logicians of *errors* and *fallacies*. For instance, ideological intuitionists accuse classical logicians of illicitly relying on the law of excluded middle—that is, for the classical logicians' own use of 'or' and 'not'. Similarly, ideological quantum logicians accuse classical logicians of illicitly relying on the distributive law—that is, for the classical logicians' own use of 'or' and 'and'. If such 'alternative' logicians were simply proposing new and more convenient or more perspicuous meanings for old words, at least as used in some theoretical contexts, such

accusations would be obviously unwarranted. Of course, the equivocation hypothesis could be extended, to imply that the 'alternative' logicians do not mean what *we* mean by words such as 'error' and 'fallacy', but applying that line to all their apparent protestations of disagreement and their extensive supporting arguments would soon suggest that they are not speaking English at all (or whatever natural language the conversation appears to be in), but some other language which we do not understand. That hypothesis blocks communication and fails to explain their linguistic behaviour in any specific way. By normal standards, we can make far more sense of them by interpreting them more or less homophonically, perhaps with a few exceptions in special theoretical contexts. Of course, their speech and writing, like almost everyone else's, is riddled with logical words; 'not', 'if', 'and', 'or', 'is', 'every', and 'some' are amongst the most commonly used words of English. Like others who have talked and corresponded with Michael Dummett, Hilary Putnam, Graham Priest, and other 'alternative' logicians, I have automatically interpreted their use of logical words in English homophonically, and never found reason to do otherwise. That was so even when the topic was 'alternative logic'. In all my experience, 'alternative' logicians use the logical words of natural language with their ordinary shared meanings, and refusing to interpret them so would be deeply uncharitable (see also Morton, 1973).

There is a connection here with Kripke's compelling critique of descriptivist views of proper names and natural kind terms (Kripke, 1972, 1980). Descriptivists tend to recycle individual differences in the beliefs associated with a given proper name or natural kind term as individual differences in its meaning, with its reference determined by descriptive fit. Similarly, the equivocation hypothesis tends to recycle individual differences in the beliefs and accepted patterns of reasoning associated with a given logical word as individual differences in its meaning, with its reference determined by descriptive fit. The equivocation hypothesis is in effect a sort of descriptivism about logical words. As Kripke and Putnam showed, descriptivism radically underestimates the social determinants of meaning. Although the implicit intention to preserve reference can fail in extreme cases, it is the standard default. Normally, we use proper names and natural kind terms as terms of a publically available language, with the references determined for them in that language. The same holds for logical terms. The equivocation hypothesis radically underestimates the social determinants of meaning (see also Williamson, 2007, pp. 121–30).

The debate over Quine's claim of equivocation in logic is also reminiscent of the debate over Kuhn and Feyerabend's claims of equivocation in natural science, which raged in the philosophy of science during the same period (Kuhn, 1962 and Feyerabend, 1970, 1975, 1978). According to Kuhn and Feyerabend, scientific revolutions involve meaning variance: old theoretical terms acquire new meanings after the revolution from the new theory, thereby making the old and new theories somehow incommensurable. In both logic and natural science, the alleged meaning variance came from differences between the theoretical systems in which the term figured, and so applied to individuals derivatively from their theoretical allegiances. Underlying Kuhn and Feyerabend's claims of meaning variance were descriptivist assumptions about meaning: reference is determined by descriptive fit. Pushback came increasingly from anti-descriptivist arguments about meaning and reference (as in

Putnam, 1973). Obviously, the creation myth of a single primordial specimen fixing the reference of a natural kind term must be taken with a pinch of salt, but that does not undermine the model of reference fixing by a fallible capacity to identify (some) specimens of the kind combined with relevant similarities and differences in their possibly unknown underlying natural properties. Such a model already undermines Kuhn and Feyerabend's descriptivist assumptions, and allows a term to refer to the same kind when used in the new theory as when used in the old theory.

Of course, the two debates did not run in perfect parallel. For example, opponents of meaning variance in the philosophy of science often invoked causal links to referents, but obviously not in the philosophy of logic. Conversely, proponents of meaning variance in the philosophy of logic often regard peaceful coexistence as a possibility for the two theories, but not in the philosophy of science. Nevertheless, in both cases, descriptivist approaches had difficulty making sense of deep theoretical disagreement, where each side is genuinely denying what the other asserts. (Newton-Smith, 1981 gives more detail on the debate over incommensurability in science.)

Do the objections to the equivocation hypothesis for logical words in a natural language extend to logical symbols in a formal language? An 'alternative' logician who presents a non-standard semantic model theory for an artificial language seems to be *stipulating* meanings for its logical symbols. That is quite different from using them as terms in a pre-existing language, with the meanings in it they already have. Might the symbol $\vee$ be sometimes used like that in quantum logic or intuitionistic logic, with a stipulated meaning subtly different from the meaning it usually has in logic (disjunction)? Several reservations need to be kept in mind.

First, we cannot assume that the new logical constants can peacefully co-exist with the old ones in a single language. For example, consider a language with two 'negation' operators, $\neg_C$, which obeys the standard rules of inference for classical negation, and $\neg_I$, which obeys the standard rules of inference for intuitionistic 'negation'; both are freely applicable to any sentence of the language. Then one can show that the two operators are logically equivalent, so $\neg_I$ *also* obeys the standard rules of inference for *classical* negation; in other words, intuitionistic negation collapses into classical negation (Popper, 1948; Harris, 1982; for discussion see Williamson, 1987/88 and Schechter, 2011).

Second, we should not take for granted that a non-standard semantic framework is even capable of defining genuine meanings for logical symbols.

Here is an example. Many intuitionists use a proof-conditional semantic framework in which the meaning of a sentence is the condition for something to be a proof of it, and the meanings of sub-sentential expressions are crafted to make the overall semantic theory compositional. Within such a framework, one can define a sentential constant $\bot$ by stipulating that nothing is to count as a proof of $\bot$. Since $\bot$ is a sentence with a well-defined meaning by the standards of such a semantic framework, we might expect to be able to add it to English. Of course, we will have to adjust it to the phonetics and syntax of English: we can do that simply by writing it as 'It bottoms', with the stipulation that it means whatever $\bot$ means. Since the symbol $\bot$ is usually called 'the *falsum*' or 'absurdity', we might expect 'It bottoms' to entail a

contradiction, say 'It is raining and not raining'. But that is unwarranted. That nothing counts as a proof of $\bot$ does not imply that $\bot$ is *false*. Perhaps there are unprovable truths. Ideological intuitionists will presumably deny that there are, but our present concern is not ideological intuitionism, just the capacity of the proof-conditional semantic framework to define genuine meanings. That framework seems to have left the meaning of $\bot$ radically vague, to the point where its English equivalent is hardly usable. That may not be total meaninglessness, but it is pretty bad.

Here is another case. Consider a logician who wants to use the standard two-valued truth-tables to define connectives, but with an alternative pair of semantic values: 'T' is reinterpreted as 'contingently true' rather than 'true' and 'F' as 'not contingently true' (false or necessarily true) rather than 'false'. In particular, the reinterpreted table for the truth-functional conditional is used to define a new connective $*$. For any formula $\alpha$, the formula $\alpha * \alpha$ always takes the value T, which now means that it is contingently true in every case. But since it is true in every case, it should be *necessarily* true, not contingently true. We could elaborate the example by setting it within a possible worlds framework; $\alpha * \alpha$ would then come out as contingently true at every world. Thus slightly reinterpreting a familiar semantic framework can turn it into garbage.

We should be wary, such examples suggest, of assuming that alternative semantic frameworks define new meanings for old logical symbols, rather than no meanings at all.

The third reservation is this. Even if 'alternative logicians' succeed in using an alternative semantic framework to stipulate new meanings for the logical symbols of a formal language, they often make clear that they intend those symbols to model the standard logical words of a natural language, and to show what the logical properties of the latter *really* are, and how they differ from the properties attributed to them by classical logicians. Typically, their arguments' aim is not merely to show that the logical words of natural language do not in fact have the properties classical logicians attribute to them; it is to show that no words *could* have those properties. Thus the stipulative aspect of such logicians' activity goes with a full-blooded rejection of classical logic.

Obviously, much of the published technical work on 'alternative logics' is carried out within an overarching classical framework and is not in disagreement with classical logic. But, in much of the more philosophical work by self-identified 'alternative logicians', the disagreement is genuine and unequivocal. That is just as anti-exceptionalism about logic predicts. Theoretical disagreements are to be expected in any fundamental science.

## 16.6 Conclusion

One could read 'The Question of Logic' as adumbrating a sort of transcendental argument for at least some principles of classical logic, an argument to show that rejecting them is not a genuine possibility for us. If the considerations of the present paper

are cogent, no such attempt could succeed. Like doing other foolish things, rejecting those principles *is* a genuine possibility for us. There is no sound transcendental argument for classical logic. That does not matter. The immanent case for classical logic is good enough. But the envisaged reading does not bring out what is most valuable in 'The Question of Logic'. Through a case study, it shows up the *glibness*, perhaps even the frivolity, of most proposals to switch to an alternative logic: their proponents have failed to think through properly what would be involved, directly and indirectly, in changing our ways of reasoning so radically. Anyone who contemplates proposing a revision of classical logical principles would be well-advised to read or reread 'The Question of Logic', and ask themselves whether they are really safe from problems analogous to those which Kripke uncovers for quantum logic.

**Acknowledgements** This paper derives from a talk to Saul Kripke and Romina Padró's 2020 CUNY seminar on the adoption problem; I thank participants for their useful questions. Many thanks also to Saul Kripke and Romina Padró for extended subsequent discussion, and to Allen Hazen for helpful correspondence (in particular, he pointed me to the passage in Whitehead, 1906). Detailed written comments from Michela Massimi and Jennifer Nagel clarified and enriched the paper in relation to the philosophy of science and cognitive psychology respectively. Two anonymous referees of an earlier version of this paper rightly encouraged me to be more explicit about various matters, including the relation of my views to Kripke's. Above all, I thank Saul Kripke himself for all his contributions to logic and philosophy, which have done so much to change the direction of inquiry for the better.

## References

Baillargeon, R. (2004). Infants' physical world. *Current Directions in Psychological Sciences,13*, 89–94.

Beall, J., & Restall, G. (2006). *Logical pluralism*. Oxford: Oxford University Press.

Besson, C. (2019). Knowledge of logical generality and the possibility of deductive reasoning. In T. Chan & A. Nes (Eds.), *Inferences and Consciousness* (pp. 172–196). London: Routledge.

Boghossian, P., & Williamson, T. (2020). *Debating the a priori*. Oxford: Oxford University Press.

Carroll, L. (1895). What the tortoise said to Achilles. *Mind, 4*, 278–280.

Dummett, M. (1975). The philosophical basis of intuitionistic logic. In H. E. Rose & J. Shepherdson (Eds.), *Logic colloquium '73* (pp. 5–40). Amsterdam: North-Holland.

Feyerabend, P. (1970). Consolations for the specialist. In I. Lakatos & A. Musgrave (Eds.), *Criticism and the growth of knowledge* (pp. 197–230). Cambridge: Cambridge University Press.

Feyerabend, P. (1975). *Against method*. London: New Left Books.

Feyerabend, P. (1978). *Science in a free society*. London: New Left Books.

Gelman, S., & Noles, N. (2011). Domains and naïve theories. *Cognitive Science, 2*, 490–502.

Ginet, C. (2001). Deciding to believe. In M. Steup (Ed.), *knowledge, truth and duty: Essays on epistemic justification, responsibility, and virtue* (pp. 63–76). New York: Oxford University Press.

Harris, J. (1982). What's so logical about the "logical" axioms? *Studia Logica, 41*, 159–171.

Hjortland, O. T. (2017). Anti-exceptionalism about logic. *Philosophical Studies, 174*, 631–658.

Humberstone, I. L. (1996). Homophony, validity, modality. In J. Copeland (Ed.), *Logic and reality: Essays on the legacy of Arthur Prior* (pp. 215–236). Oxford: Oxford University Press.

Kripke, S. (1963). Semantical considerations on modal logic. *Acta Philosophica Fennica, 16*, 83–94. Reprinted in Linsky, 1971, 63–72.

Kripke, S. (1972). Naming and necessity. In D. Davidson & G. Harman (Eds.), *Semantics of natural language* (pp. 253–355, 763–769). Dordrecht: Reidel. Reprinted with additions as Kripke, 1980.
Kripke, S. (1975). Outline of a theory of truth. *Journal of Philosophy, 72*(19), 690–716.
Kripke, S. (1980). *Naming and necessity*. Oxford: Blackwell.
Kripke, S. (2013). *Reference and existence: The John Locke lectures*. Oxford: Oxford University Press.
Kripke, S. (2020). The question of logic. MS.
Kubricht, J., Holyoak, K. & Lu, H. (2017). Intuitive physics: current research and controversies. *Trends in Cognitive Sciences, 21*, 749–759.
Kuhn, T. (1962). *The structure of scientific revolutions*. Chicago: University of Chicago Press.
Linsky, L. (Ed.). (1971). *Reference and modality*. Oxford: Oxford University Press.
Marcus, R. B. (1946). A functional calculus of first order based on strict implication. *Journal of Symbolic Logic, 11*, 1–16.
McCarty, C. (2008). Completeness and incompleteness for intuitionistic logic. *Journal of Symbolic Logic, 73*, 1315–1327.
Morton, A. (1973). Denying the doctrine and changing the subject. *Journal of Philosophy, 70*, 503–510.
Newton-Smith, W. H. (1981). *The rationality of science*. London: Routledge.
Popper, K. (1948). On the theory of deduction, part II: The definitions of classical and intuitionist negation. *Indagationes Mathematicae, 10*, 111–120.
Putnam, H. (1968). Is logic empirical? In R. Cohen & M. Wartofsky (Eds.), *Boston studies in philosophy of science* (Vol. 6, pp. 216–241). Dordrecht: Reidel.
Putnam, H. (1973). Explanation and reference. In G. Pearce & P. Maynard (Eds.), *Conceptual change* (pp. 199–221). Dordrecht: Reidel.
Putnam, H. (2005). A philosopher looks at quantum mechanics (again). *British Journal for the Philosophy of Science, 56*, 615–634.
Quine, W. V. O. (1936). Truth by convention. In O. H. Lee (Ed.), *Philosophical essays for A. N. Whitehead* (pp. 90–124). New York: Longmans.
Quine, W. V. O. (1951). Two dogmas of empiricism. *Philosophical Review, 60*, 20–43.
Quine, W. V. O. (1970). *Philosophy of logic*. Englewood Cliffs: Prentice Hall.
Rott, H. (2017). Negative doxastic voluntarism and the concept of belief. *Synthese, 194*, 2695–2720.
Schechter, J. (2011). Juxtaposition: A new way to combine logics. *Review of Symbolic Logic, 4*, 560-606.
Sliwa, P. (2015). Understanding and knowing. *Proceedings of the Aristotelian Society, 115*, 57–74.
Smith, B. & Casati, R. (1994). Naïve physics. *Philosophical Quarterly, 7*, 227–247.
Spelke, E., & Kinzler, K. (2007). Core knowledge. *Developmental Science, 10*, 89–96.
Tarski, A. (1935). Der Wahrheitsbegriff in den formalisierten Sprachen. *Studia Philosophica, 1*, 261–405. Reprinted as Woodger, J. H. (Trans.). (1983). The concept of truth in formalized languages. In *Logic, semantics, metamathematics* (2nd ed., pp. 153–278). Indianapolis: Hackett.
van Fraassen, B. (1980). *The scientific image*. New York: Oxford University Press.
van Fraassen, B. (2002). *The empirical stance*. New Haven: Yale University Press.
Weatherson, B. (2008). Deontology and Descartes' demon. *Journal of Philosophy, 105*, 540–569.
Whitehead, A. N. (1906). On mathematical concepts of the material world. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences, 205*, 465–525.
Williamson, T. (1987/88). Equivocation and existence. *Proceedings of the Aristotelian Society, 88*, 109–127.
Williamson, T. (1994). *Vagueness*. London: Routledge.
Williamson, T. (2000). *Knowledge and its limits*. Oxford: Oxford University Press.
Williamson, T. (2007). *The philosophy of philosophy*. Oxford: Blackwell.
Williamson, T. (2013). *Modal logic as metaphysics*. Oxford: Oxford University Press.
Williamson, T. (2018). Alternative logics and applied mathematics. *Philosophical Issues, 28*, 399–424.

Williamson, T. (2020). *Suppose and tell: The semantics and heuristics of conditionals*. Oxford: Oxford University Press.
Williamson, T. (2024). Knowledge, credence, and strength of belief. In A. K. Flowerree & B. Reed (Eds.), *Expansive epistemology: Norms, action, and the social world*. London: Routledge.

# Appendix A
# Saul Kripke: Complete Bibliography


Kripke, S. A. (1959a). A completeness theorem in modal logic. *Journal of Symbolic Logic, 24*(1), 1–14.

Kripke, S. A. (1959b). Semantical analysis of modal logic (abstract). *Journal of Symbolic Logic, 24*(4), 323–324.

Kripke, S. A. (1959c). The problem of entailment (abstract). *Journal of Symbolic Logic, 24*(4), 324.

Kripke, S. A. (1959d). Distinguished constituents (abstract). *Journal of Symbolic Logic, 24*(4), 323.

Kripke, S. A. (1962a). 'Flexible' predicates of formal number theory. *Proceedings of the American Mathematical Society, 13*(4), 647–650.

Kripke, S. A. (1962b). The undecidability of monadic modal quantification theory. *Zeitschrift für mathematische Logik und Grundlagen der Mathematik, 8*, 113–116.

Kripke, S. A. (1963a). Semantical considerations on modal logic. *Acta Philosophica Fennica, 16*, 83–94.

Kripke, S. A. (1963b). Semantical analysis of modal logic I. Normal modal propositional calculi. *Zeitschrift für mathematische Logik und Grundlagen der Mathematik, 9*, 67–96.

Kripke, S. A. (1964a). Transfinite recursions on admissible ordinals, I (abstract). *Journal of Symbolic Logic, 29*(3), 161.

Kripke, S. A. (1964b). Transfinite recursions on admissible ordinals, II (abstract). *Journal of Symbolic Logic, 29*(3), 161–162.

Kripke, S. A. (1964c). Admissible ordinals and the analytic hierarchy (abstract). *Journal of Symbolic Logic, 29*(3), 162.

Kripke, S. A. (1965a). Semantical analysis of modal logic II. Non-normal modal propositional calculi. In J. W. Addison, L. Henkin, & A. Tarski (Eds.), *The theory of models, proceedings of the 1963 international symposium at Berkeley* (pp. 206–220). Amsterdam: North Holland Publishing Company.

Kripke, S. A. (1965b). Semantical analysis of intuitionistic logic I. In J. N. Crossley & M. A. E. Dummett (Eds.), *Formal systems and recursive functions, proceedings of the eighth logic colloquium at Oxford, July, 1963* (pp. 92–130). Amsterdam: North Holland Publishing Company.

Kripke, S. A. (1967a). Review of E. J. Lemmon, 'Algebraic semantics for modal logic II' (Journal of Symbolic Logic, 31 (1966), pp. 191–218). *Mathematical Reviews*, *34*(5), 1022.

Kripke, S. A. (1967b). Review of E. J. Lemmon, 'Algebraic semantics for modal logic I' (Journal of Symbolic Logic, 31 (1966), pp. 46–65). *Mathematical Reviews, 34*(5), 1021–1022.

Kripke, S. A. (1967c). Transfinite recursion, constructible sets, and analogues of cardinals. In *Summaries of talks prepared in connection with the summer institute on axiomatic set theory* (pp. IV01–12). UCLA (Los Angeles): American Mathematical Society.


© The Editor(s) (if applicable) and The Author(s), under exclusive license to Springer Nature Switzerland AG 2024

Y. Weiss and R. Birman (eds.), *Saul Kripke on Modal Logic*, Outstanding Contributions to Logic 30, https://doi.org/10.1007/978-3-031-57635-5

Kripke, S. A. (1967d). On the application of Boolean-valued models to solutions of problems in Boolean algebra. In *Summaries of talks prepared in connection with the summer institute on axiomatic set theory* (pp. IVT1–7). UCLA (Los Angeles): American Mathematical Society.
Kripke, S. A. (1967e). An extension of a theorem of Gaifman-Hales-Solovay. *Fundamenta Mathematicae, 61*, 29–32.
Kripke, S. A. (1971). Identity and necessity. In M. K. Munitz (Ed.), *Identity and individuation* (pp. 135–164). New York: New York University Press.
Kripke, S. A. (1972). Naming and necessity. In D. Davidson & G. Harman (Eds.), *Semantics of natural language* (2nd ed., pp. 253–355, 763–769). Dordrecht: D. Reidel Publishing Company.
Kripke, S. A. (1975). Outline of a theory of truth. *Journal of Philosophy, 72*(19), 690–716.
Kripke, S. A. (1976a). A theory of truth I. Preliminary report (abstract). *Journal of Symbolic Logic, 41*(2), 556.
Kripke, S. A. (1976b). A theory of truth II. Preliminary report (abstract). *Journal of Symbolic Logic, 41*(2), 556–557.
Kripke, S. A. (1976c). Is there a problem about substitutional quantification? In G. Evans & J. McDowell (Eds.), *Truth and meaning* (pp. 325–419). London: Oxford University Press.
Kripke, S. A. (1977). Speaker's reference and semantic reference. *Midwest Studies in Philosophy, 2*(1), 255–276.
Kripke, S. A. (1979). A puzzle about belief. In A. Margalit (Ed.), *Meaning and use* (pp. 239–283). Dordrecht: D. Reidel Publishing Company.
Kripke, S. A. (1980). *Naming and necessity*. Cambridge: Harvard University Press.
Kripke, S. A. (1981). Wittgenstein on rules and private language: An elementary exposition. In I. Block (Ed.), *Perspectives on the philosophy of Wittgenstein* (pp. 238–312). Oxford: Basil Blackwell.
Kripke, S. A. (1982). *Wittgenstein on rules and private language*. Cambridge: Harvard University Press.
Kripke, S. A. (1983). Review of Kit Fine, 'Failures of the interpolation lemma in quantified modal logic' (Journal of Symbolic Logic, 44 (1979), pp. 201–206). *Journal of Symbolic Logic, 48*(2), 486–488.
Kripke, S. A. (1985). Review of Kit Fine, 'Model theory for modal logic. Parts I-III' (Journal of Philosophical Logic, 7 (1978), pp. 125–156, 277–306; and 10 (1981), pp. 293–307). *Journal of Symbolic Logic, 50*(4), 1083–1093.
Kripke, S. A. (1986). A problem in the theory of reference: The linguistic division of labor and the social character of naming. In *Philosophy and culture (proceedings of the XVIIth world congress of philosophy)* (pp. 241–247). Montreal: Editions du Beffroi, Editions Montmorency.
Kripke, S. A. (1992). Individual concepts: Their logic, philosophy, and some of their uses (abstract). *Proceedings and Addresses of the American Philosophical Association, 66*(2), 70–73.
Kripke, S. A. (2005). Russell's notion of scope. *Mind, 114*(456), 1005–1037.
Kripke, S. A. (2008). Frege's theory of sense and reference: Some exegetical notes. *Theoria, 74*(3), 181–218.
Kripke, S. A. (2009a). The collapse of the Hilbert Program (abstract). *Bulletin of Symbolic Logic, 15*(2), 229–231.
Kripke, S. A. (2009b). Presupposition and anaphora: Remarks on the formulation of the projection problem. *Linguistic Inquiry, 40*(3), 367–386.
Kripke, S. A. (2011). *Philosophical troubles: Collected papers* (Vol. 1). Oxford: Oxford University Press.
Kripke, S. A. (2013a). The Church-Turing 'Thesis' as a special corollary of Gödel's completeness theorem. In B. J. Copeland, C. Posy, & O. Shagrir (Eds.), *Computability: Turing, Gödel, Church, and beyond* (pp. 77–104). Cambridge: MIT Press.
Kripke, S. A. (2013b). *Reference and existence: The John Locke lectures*. Oxford: Oxford University Press.
Kripke, S. A. (2014a). Fregean quantification theory. *Journal of Philosophical Logic, 43*(5), 879–881.

Kripke, S. A. (2014b). The road to Gödel. In J. Berg (Ed.), *Naming, necessity, and more: Explorations in the philosophical work of Saul Kripke* (pp. 223–241). New York: Palgrave Macmillan.
Kripke, S. A. (2015). Yet another dogma of empiricism. *Philosophy and Phenomenological Research, 91*(2), 381–385.
Kripke, S. A. (2017a). Quantified modality and essentialism. *Noûs, 51*(2), 221–234.
Kripke, S. A. (2017b). Quantified modal logic and Quine's critique: Some further observations. *Noûs, 51*(2), 235–237.
Kripke, S. A. (2017c). 'And' and 'but': A note. *Thought, 6*(2), 102–105.
Kripke, S. A. (2017d). History and idealism: The theory of R. G. Collingwood. *Collingwood and British Idealism Studies, 23*(1), 9–29.
Kripke, S. A. (2019a). Ungroundedness in Tarskian languages. *Journal of Philosophical Logic, 48*(3), 603–609.
Kripke, S. A. (2019b). Free choice sequences: A temporal interpretation compatible with acceptance of classical mathematics. *Indagationes Mathematicae, 30*(3), 492–499.
Kripke, S. A. (2022a). A proof of Gamma. In K. Bimbó (Ed.), *Relevance logics and other tools for reasoning. Essays in honor of J. Michael Dunn* (pp. 261–265). London: College Publications.
Kripke, S. A. (2022b). Mathematical incompleteness results in first-order Peano arithmetic: A revisionist view of the early history. *History and Philosophy of Logic, 43*(2), 175–182.
Kripke, S. A. (2022c). The collapse of the Hilbert Program: A variation on the Gödelian theme. *Bulletin of Symbolic Logic, 28*(3), 413–426.
Kripke, S. A. (2023a). Wittgenstein, Russell, and our concept of the natural numbers. In C. Posy & Y. Ben-Menahem (Eds.), *Mathematical knowledge, objects and applications: Essays in memory of Mark Steiner* (pp. 137–155). Cham: Springer.
Kripke, S. A. (2023b). Gödel's theorem and direct self-reference. *Review of Symbolic Logic, 16*(2), 650–654.
Kripke, S. A. (2024). The question of logic. *Mind, 133*(529), 1–36.
Kripke, S. A. (in press-a). Individual concepts: Their logic, philosophy, and some of their uses. In Y. Weiss & R. Birman (Eds.), *Saul Kripke on modal logic*. Cham: Springer.
Kripke, S. A. (in press-b). A letter from Kripke to Lewis. In Y. Weiss & R. Birman (Eds.), *Saul Kripke on modal logic*. Cham: Springer.
Kripke, S. A., & Kochen, S. (1982). Non-standard models of Peano arithmetic. *L'Enseignement Mathématique, 28*(1–2), 211–231.
Kripke, S. A., & Pour-El, M. B. (1967a). Deduction-preserving 'recursive isomorphisms' between theories (research announcement). *Bulletin of the American Mathematical Society, 73*, 145–148.
Kripke, S. A., & Pour-El, M. B. (1967b). Deduction-preserving 'recursive isomorphisms' between theories. *Fundamenta Mathematicae, 61*, 141–163.
Kripke, S. A., Marcus, R. B., Quine, W. V., McCarthy, T., & Føllesdal, D. (1963). Discussion (colloquium on 'modalities and intensional languages'). In M. W. Wartofsky (Ed.), *Colloquium for the philosophy of science, Boston, 1961–2* (Vol. 1, pp. 105–116). Dordrecht: D. Reidel Publishing Company.
Kripke, S. A., Davidson, S., Kaplan, D., Quine, W. V., Partee, B., Dummett, M. A. E., ..., & Parsons, C. (1974a). First general discussion session (proceedings of a conference on 'language, intentionality, and translation-theory'). *Synthese, 27*(3/4), 471–508.
Kripke, S. A., Harman, G., Quine, W. V., Lewis, D., Dummett, M. A. E., & Partee, B. (1974b). Second general discussion session (proceedings of a conference on 'language, intentionality, and translation-theory'). *Synthese, 27*(3/4), 509–521.